Detecting Compromised Accounts on Social Networks using Anomalous User Behaviors

Aarti Varpe[1], Manjushri Mahajan[2], Pournima More[3]
[1]M.tech Student, [2]Assistant Professor (Guide), [3]Assistant Professor (Co-guide)
Department of Computer Engineering, G.H.R.C.E.M, Wagholi, Savitribai Phule Pune University, Maharashtra, India

Abstract - On web account Compromising is a serious threat to users of Online Social Networks (OSNs). Instead of analyzing user profile contents or message contents, present system seek to uncover the behavioral anomaly of compromised accounts by using their legitimate owners' history social activity patterns, which can be observed in a lightweight manner. To understand user all activity OSN provide variety of online features for their users to engage in, such as building connections, sending messages, uploading photos, browsing friends' latest updates, etc. However, how a user involves in each activity is completely driven by personal interests and social habits. While a user tends to conform to its social patterns, a hacker of the user account who knows little about the user's behavior habit is likely to diverge from the patterns. This paper presents a novel method to detect Account Compromisation as and when it happens by profiling online social behaviors. System uses Naïve baye's account to analyze tweets are positive or negative.

Keywords **-** Online social networks, cybercrime, network security

I. INTRODUCTION

Account Compromising is a serious threat to users of Online Social Networks (OSNs). While relentless spammers exploit the established trust relationships between account owners and their friends to efficiently spread malicious spam, timely detection of compromised accounts is quite challenging due to the well-established trust relationship between the service providers, account owners, and their friends. Instead of analyzing user profile contents or message contents, present system seek to uncover the behavioral anomaly of compromised accounts by using their legitimate owners' history social activity patterns, which can be observed in a lightweight manner. To better serve users' various social communication needs, OSNs provide a great variety of online features for their users to engage in, such as building connections, sending messages, uploading photos, browsing friends' latest updates, etc. However, how a user involves in each activity is completely driven by personal interests and social habits. As a result, the interaction patterns with a number of OSN activities tend to be divergent across a large set of users. While a user tends to conform to its social patterns, a hacker of the user account who knows little about the user's behavior habit is likely to diverge from the patterns. This paper presents a novel method to detect Account Compromisation as and when it happens by profiling online social behaviors.

II. BACKGROUND

Previous research on spamming account detection mostly cannot distinguish compromised accounts from Sybil accounts, with only one recent study by Egele et al. features compromised accounts detection. Existing approaches involve account profile analysis and message content analysis (e.g. embedded URL analysis and message clustering). However, account profile analysis is hardly applicable for detecting compromised accounts, because their profiles are the original common users' information which is likely to remain intact by spammers. Malicious parties exploit the well-established connections and trust relationships between the legitimate account owners and their friends, and efficiently distribute spam ads, phishing links, or malware, while avoiding being blocked by the service providers. Major OSNs today employ IP relocation logging to battle against account Compromisation. However, this approach is known to suffer from low detection granularity and high false positive.

III. REVIEW OF LITERATURE

They created a labeled collection with users classified as spammers or non-spammers. They provided a characterization of the users of this labeled collection, bringing to the light several attributes useful to differentiate spammers and non-spammers. They leverage our characterization study towards a spammer detection mechanism. Using a classification technique, They were able to correctly identify a significant fraction of the spammers while incurring in a negligible fraction of misclassification of legitimate users[1].

In this paper, overall research goal is to investigate techniques and develop effective tools for automatically detecting and filtering spammers who target social systems[2].

This paper aims to identify single spam bots, as well as large-scale campaigns. present system also showed how our techniques help to detect spam profiles even when they do not contact a honey-profile[3].

In this paper, they describe our work to provide online spam filtering for social networks. They use text shingling and URL comparison to incrementally reconstruct spam messages into campaigns, which are then identified by a trained classifier[4].

In this paper, they describe our work on detecting and characterizing spam campaigns performed using asynchronous wall messages on the Face book social network. They analyze a large dataset composed of over 187 million wall posts written to the profile walls of 3.5 million Face book users[5].

In this paper, they propose a new suspicious URL detection system for Twitter, warningbird. Unlike the previous systems, warning bird is robust when protecting against conditional redirection, because it does not rely on the features of malicious landing pages that may not be reachable. Instead, it focuses on the correlations of multiple redirect chains that share redirection servers[6].

In particular, they showed that email spam provides little insight into the properties of Twitter spammers, while the reverse is also true. They explored the distinctions between email and Twitter spam, including the overlap of spam features, the persistence of features over time, and the abuse of generic redirectors and public web hosting[7].

In this paper, they presented a novel approach to detect compromised accounts in social networks. More precisely, they developed statistical models to characterize the behavior of social network users, and they used anomaly detection techniques to identify sudden changes in their behavior[8].

This paper aims to answer the question: Are social links valid indicators of real user interaction? To do this, they gathered extensive data from crawls of the Face book social network, including social and interaction statistics on more than 10 million users. They show that interaction activity on Face book is significantly skewed towards a small portion of each user's social links. This finding casts doubt on the assumption that all social links imply equally meaningful friend relationships[9].

In order to identify influential's on Twitter, they have ranked users by the number of followers and by Page Rank and found two rankings to be similar. If they rank by the number of re-tweets, then the ranking differs from the previous two rankings, indicating a gap in influence inferred from the number of followers and that from the popularity of one's tweets. Ranking by re-tweets exposes the influence of other media in a novel perspective[10].

IV. PROPOSED METHODOLOGY

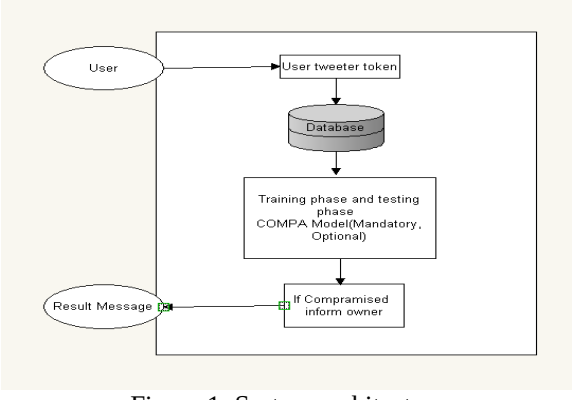


Figure 1: System architecture

System Overview - The system users tweeter dataset. In this paper present system present COMPA, the first detection system designed to identify compromised social network accounts. COMPA is based on a simple observation: social network users develop habits over time, and these habits are fairly stable. Conversely, if the account falls under the control of an adversary, the messages that the attacker sends will likely show anomalies compared to the typical behavior of the user.To detect account compromises, COMPA builds a behavioral profile for social network accounts, based on the messages sent by the account in the past. Every time a new message is generated, the message is compared against this behavioral profile. If the message significantly deviates from the learned behavioral profile, COMPA flags it as a possible compromise. Once our system has obtained the message stream for a user, we use this information to build the corresponding behavioral profile. the system extracts a set of feature values from each message, and then, for each feature, trains a model. Each of these models captures a characteristic feature of a message, such as the time the message was sent, or the application that was used to generate it. Given the behavioral profile for a user, we can assess to what extent a new message corresponds to the expected behavior. To this end, we compute the anomaly score for a message with regard to the user's established profile. The anomaly score is computed by extracting the feature values for the new message, and then comparing these feature values to the corresponding feature models. Each model produces 0 and 1where 0 denotes perfectly normal and 1 indicates that the feature is highly anomalous. The anomaly score for a message is then calculated by composing the results for all individual models. Model characteristics are Time (Hour of Day), Message Source, Message Text (Language), Links in Messages, Message Topic, URL Similarity. Naive baye's algorithm used for tweets classification that all the tweets of that account is positive or negative.

Advantages:

1. This system uses COMPA that the first system designed to detect compromised social network accounts.
2. This system can reliably detect compromised account that affect high profile accounts.

Hardware Requirements:

Processor	Pentium iv/intel i3 core
Speed	1.1 GHz
RAM	2GB
Hard Disk	50GB
Keyboard	Standard Keyboard
Mouse	Two or Three Button
Monitor	Led Monitor

Software Requirements:

Operating System	Windows Xp/7
Programming Language	Java/J2ee
Software Version	Jdk 1.7 Or Above
Tools	Eclipse
Front End	Jsp
Database	Mysql

V. MATHEMATIAL MODEL

The feature fv for the analyzed model is first extracted from the message. If Mf contains a tuple with fv as a first element, then the tuple <fv; ci >is extracted from Mf . If there is no tuple in Mf with fv as a first value, the message is considered anomalous. The procedure terminates here and an anomaly score of 1 is returned. Each feature model is represented as a set Mf . Each element of

M(f) is a tuple <fv; ci.>----------1

$$M(f)=\sum_{i=1}^{M(f)} C(i)/N \quad \text{----------2}$$

C(i) is, for each tuple in Mf , the second element of the tuple. If c is greater or equal than _Mf , the message is considered to comply with the learned behavioral profile for that feature, and an anomaly score of 0 is returned.

VI. ALGORITHM

a) Preprocessing:

1. Stop word Removal-This technique remove stop words like is, are, they, but etc.
2. Tokenization-This technique remove Special character and images.
3. Stemming remove suffix and prefix and Find Original word for e.g.- Played play

b) Naive Baye's:

Naive Baye's: This algorithm is used classify posts is positive or negative

Input: Post

Output: Predicated class of posts.

Working:

Step 1: Take posts

Step 2: Preprocess the posts

Step 3: Pass to naive Bayes class.

Step 4: Get positive and negative score according to specify its dictionary.

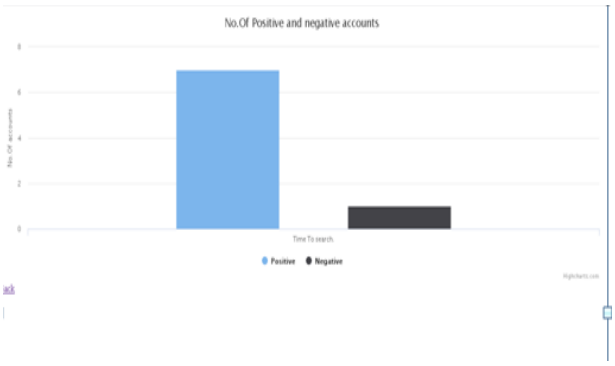
Step 5: Get max score and declare as positive or negative.

Step 6: Predicated class of all posts and analyze compromised account

VII. RESULT AND DISCUSSION

Table 1: demonstrated the No.of compromised account that is negative and No.of Positive account.

Sr.no		No.of Accounts
1	Positive	7
2	Negative	1



Graph 1:X-axis Positive or negative Y-axis No.of accounts.

Graph 01: showed a pictorial representation of Proposed system Positive and negative account analyze with the parameter considering Model characteristics are Time (Hour of Day), Message Source, Message Text (Language), Links in Messages, Message Topic, URL Similarity

VIII. CONCLUSION

A social behavioral profile for individual OSN users to characterize their behavioral patterns is proposed and built. Based on the characterized social behavioral profiles, system is able to distinguish a user from others, which can be easily employed for compromised account detection. The results show that our approach can reliably detect compromises affecting high profile social network accounts, and can detect compromises of regular accounts. Naive Baye's algorithm will classify tweets Positive or negative of user account.

IX. REFERENCES

[1]. F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on twitter," in Proc. Conf. Email Anti-Spam, 2010, vol. 6, p. 12.

[2]. K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: Social honeypots + machine learning," in Proc. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2010, pp. 435–442.

[3]. G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in Proc. 26th Annu. Comput. Security Appl. Conf., 2010, pp. 1–9.

[4]. H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary, "Towards online spam filtering in social networks," in Proc. Symp. Netw. Distrib. Syst. Security, 2012.

[5]. H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao, "Detecting and characterizing social spam campaigns," in Proc. 10th ACM SIGCOMM Conf. Internet Meas., 2010, pp. 35–47.

[6]. S. Lee and J. Kim, "WarningBird: Detecting suspicious URLs in twitter stream," in Proc. Symp. Netw. Distrib. Syst. Security, 2012.

[7]. K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time URL spam filtering service," in Proc. IEEE Symp. Security Privacy, 2011, pp. 447–462.

[8]. M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "COMPA: Detecting compromised accounts on social networks," presented at the Network and Distributed System Security Symp., San Diego, CA, USA Feb. 2013.

[9]. C. Wilson, B. Boe, A. Sala, K. Puttaswamy, and B. Zhao, "User interactions in social networks and their implications," in Proc. 26th Annu. Comput. Security Appl. Conf., 2010, pp. 11–20.

[10]. H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in Proc. 19th Int. Conf. World Wide Web, 2010, pp. 591–600.

[11]. S. Yardi, D. Romero, G. Schoenebeck, and D. Boyd, "Detecting spam in a twitter network," First Monday, vol. 15, no. 1, 2010.

[12]. Z. Cai and C. Jermaine, "The latent community model for detecting sybils in social networks," in Proc. Symp. Netw. Distrib. Syst. Security, 2012, pp. 563–578.

[13]. C. Ghiossi. (2010). Explaining Facebook's spam prevention systems [Online]. Available: http://blog.facebook.com/blog.php? post=403200567130

[14]. Twitter. (2010). The twitter rules [Online]. Available: http:// support.twitter.com/entries/18311-the-twitter-rules
[15]. . Xu, F. Zhang, and S. Zhu, "Toward worm detection in online social networks," in Proc. 26th Annu. Comput. Security Appl. Conf., 2010, pp. 11–20.
[16]. J. Baltazar, J. Costoya, and R. Flores, "KOOBFACE: The largest web 2.0 botnet explained," 2009.
[17]. Z. Chu, S. Giannivecchio, H. Wang, and S. Jajodia, "Who is tweeting on Twitter: Human, bot, or cyborg?" in Proc. 26th Annu. Comput. Security Appl. Conf., 2010, pp. 21–30.
[18]. C. Yang, R. Harkreader, and G. Gu, "Die free or live hard? Empirical evaluation and new design for fighting evolving twitter spammers," in Proc. Symp. Recent Adv. Intrusion Detection, 2011, pp. 318–337.
[19]. G. Stringhini, P. Mourlanne, G. Jacob, M. Egele, C. Kruegel, and G. Vigna, "EvilCohort: Detecting communities of malicious accounts on online services," in Proc. USENIX Security Symp., 2015.
[20]. Q. Cao, x. Yang, J. Yu, and C. Palow, "Uncovering large groups of active malicious accounts in online social networks,"