



Detecting Compromised Accounts on Social Networks

¹KURELLA MOUNIKA, ² Dr CH VENKATRAMANA REDDY

M.Tech, Department of Computer Science and Engineering,

Associate Professor, Department of CSE

Swami Ramananda Tirtha Institute of Science & Technology, Nalgonda.

ABSTRACT

Compromising social network accounts has become a profitable course of action for cybercriminals. By hijacking control of a popular media or business account, attackers can distribute their malicious messages or disseminate fake information to a large user base. The impacts of these incidents range from a tarnished reputation to multi-billion dollar monetary losses on financial markets. In our previous work, we demonstrated how we can detect large-scale compromises (i.e., so-called campaigns) of regular online social network users. In this work, we show how we can use similar techniques to identify compromises of individual high-profile accounts. High-profile accounts frequently have one characteristic that makes this detection reliable – they show consistent behavior over time. We show that our

system, were it deployed, would have been able to detect and prevent three real-world attacks against popular companies and news agencies. Furthermore, our system, in contrast to popular media, would not have fallen for a staged compromise instigated by a US restaurant chain for publicity reasons.

1. INTRODUCTION

Online social networks, such as Face book and Twitter, have become one of the main media to stay in touch with the rest of the world. Celebrities use them to communicate with their fan base, corporations take advantage of them to promote their brands and have a direct connection to their customers, while news agencies leverage social networks to distribute breaking news. Regular users make pervasive use of social networks too, to stay in touch with their

friends or colleagues and share content that they find interesting.

Over time, social network users build trust relationships with the accounts they follow. This trust can develop for a variety of reasons. For example, the user might know the owner of the trusted account in person or the account might be operated by an entity commonly considered as trustworthy, such as a popular news agency. Unfortunately, should the control over an account fall into the hands of a cyber criminal, he can easily exploit this trust to further his own malicious agenda. Previous research showed that using compromised accounts to spread malicious content is advantageous to cyber criminals, because social network users are more likely to react to messages coming from accounts they trust.

2. Literature Survey

Detecting Spammers on Twitter

In this paper we consider the problem of detecting spammers on Twitter. We first collected a large dataset of Twitter that includes more than 54 million users, 1.9 billion links, and almost 1.8 billion tweets. Using tweets related to three famous

trending topics from 2009, we construct a large labeled collection of users, manually classified into spammers and non-spammers. We then identify a number of characteristics related to tweet content and user social behavior, which could potentially be used to detect spammers. We used these characteristics as attributes of machine learning process for classifying users as either spammers or non spammers. Our strategy succeeds at detecting much of the spammers while only a small percentage of non-spammers are misclassified. Approximately 70 % of spammers and 96% of non-spammers were correctly classified. Our results also highlight the most important attributes for spam detection on Twitter.

Uncovering social spammers: social honeypots + machine learning

We propose and evaluate a honeypot-based approach for uncovering social spammers in online social systems. Two of the key components of the proposed approach are: (1) The deployment of social honeypots for harvesting deceptive spam profiles from social networking communities; and (2) Statistical analysis of the properties of these spam profiles for creating spam classifiers to actively filter out existing and new



spammers. We describe the conceptual framework and design considerations of the proposed approach, and we present concrete observations from the deployment of social honeypots in MySpace and Twitter. We find that the deployed social honeypots identify social spammers with low false positive rates and that the harvested spam data contains signals that are strongly correlated with observable profile features (e.g., content, friend information, posting patterns, etc.). Based on these profile features, we develop machine learning based classifiers for identifying previously unknown spammers with high precision and a low rate of false positives.

Detecting Spammers on Social Networks

In this paper, we analyze to which extent spam has entered social networks. More precisely, we analyze how spammers who target social networking sites operate. To collect the data about spamming activity, we created a large and diverse set of "honey-profiles" on three large social networking sites, and logged the kind of contacts and messages that they received. We then analyzed the collected data and identified anomalous behavior of users who contacted

our profiles. Based on the analysis of this behavior, we developed techniques to detect spammers in social networks, and we aggregated their messages in large spam campaigns. Our results show that it is possible to automatically identify the accounts used by spammers, and our analysis was used for take-down efforts in a real-world social network. More precisely, during this study, we collaborated with Twitter and correctly detected and deleted 15,857 spam profiles.

Towards Online Spam Filtering in Social Networks

In this paper, a supervised machine learning based solution is proposed for an effective spammer detection. The main procedure of the work is: first, collect a dataset from Sina Weibo including 30,116 users and more than 16 million messages. Then, construct a labeled dataset of users and manually classify users into spammers and non-spammers. Afterwards, extract a set of feature from message content and users' social behavior, and apply into SVM (Support Vector Machines) based spammer detection algorithm. The experiment shows that the proposed solution is capable to

provide excellent performance with true positive rate of spammers and non-spammers reaching 99.1% and 99.9% respectively.

Compa: Detecting compromised accounts on social networks

In this paper, we present a novel approach to detect compromised user accounts in social networks, and we apply it to two popular social networking sites, Twitter and Face book. Our approach uses a composition of statistical modeling and anomaly detection to identify accounts that experience a sudden change in behavior. Since behavior changes can also be due to benign reasons (e.g., a user could switch her preferred client application or post updates at an unusual time), it is necessary to derive a way to distinguish between malicious and legitimate changes. To this end, we look for groups of accounts that all experience similar changes within a short period of time, assuming that these changes are the result of a malicious campaign that is unfolding. We developed a tool, called COMPA, that implements our approach, and we ran it on a large-scale dataset of more than 1.4 billion publicly-available Twitter

messages, as well as on a dataset of 106 million Face book messages. COMPA was able to identify compromised accounts on both social networks with high precision. 1

@spam: the underground on 140 characters or less

In this work we present a characterization of spam on Twitter. We find that 8% of 25 million URLs posted to the site point to phishing, malware, and scams listed on popular blacklists. We analyze the accounts that send spam and find evidence that it originates from previously legitimate accounts that have been compromised and are now being puppeteer by spammers. Using click through data, we analyze spammers' use of features unique to Twitter and the degree that they affect the success of spam. We find that Twitter is a highly successful platform for coercing users to visit spam pages, with a click through rate of 0.13%, compared to much lower rates previously reported for email spam. We group spam URLs into campaigns and identify trends that uniquely distinguish phishing, malware, and spam, to gain an

insight into the underlying techniques used to attract users.

Given the absence of spam filtering on Twitter, we examine whether the use of URL blacklists would help to significantly stem the spread of Twitter spam. Our results indicate that blacklists are too slow at identifying new threats, allowing more than 90% of visitors to view a page before it becomes blacklisted. We also find that even if blacklist delays were reduced, the use by spammers of URL shortening services for obfuscation negates the potential gains unless tools that use blacklists develop more sophisticated spam filtering.

3. OVERVIEW OF THE SYSTEM

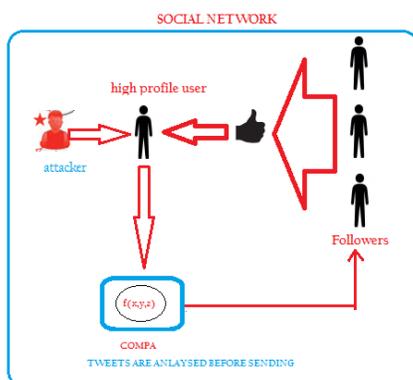


Fig 3.1 System Architecture

EXISTING SYSTEM:

Grier et al. studied the behavior of compromised accounts on Twitter by entering the credentials of an account they controlled on a phishing campaign site.

Gao et al- developed a clustering approach to detect spam wall posts on Face book. They also attempted to determine whether an account that sent a spam post was compromised. To this end, the authors look at the wall post history of spam accounts. However, the classification is very simple. When an account received benign wall post from one of their connections (friends), they automatically considered that account as being legitimate but compromised. The problem with this technique is that previous work showed that spam victims occasionally send messages to these spam accounts. This would cause their approach to detect legitimate accounts as compromised.

Disadvantages of Existing System:

This approach does not scale as it requires identifying and joining each new phishing campaign. Also, this approach is limited to phishing campaigns.

Gao et al. Developed system needs to know whether an account has sent spam before it can classify it as fake or compromised.



PROPOSED SYSTEM:

We in this paper, we 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. A typical social network user, for example, might consistently check her posts in the morning from her phone, and during the lunch break from her desktop computer. Furthermore, interaction will likely be limited to a moderate number of social network contacts (i.e., friends). 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.

We show that COMPA can reliably detect compromises that affect high profile accounts. Since the behavior of these accounts is very consistent, false positives are minimal.

To detect large-scale compromises, we propose to group similar messages together and apply COMPA to them, to assess how many of those messages violate their

accounts' behavioral profile. This grouping accounts for the fact that regular social network accounts show a more variable behavior compared to high profile ones, and allows us to keep false positives low.

Advantages of Proposed System:

COMPA uses statistical models to characterize the behavior of social network users, and leverages anomaly detection techniques to identify sudden changes in their behavior.

The results show that our approach can reliably detect compromises affecting high-profile social network accounts, and can detect compromises of regular accounts, whose behavior is typically more variable, by aggregating together similar malicious messages.

Our system, on the other hand, detects compromised accounts also when they are not involved in spam campaigns.

MODULES

User:

OSN System Construction Module

In the first module, we develop the Online Social Networking (OSN) system module.

We build up the system with the feature of Online Social Networking. Where, this module is used for new user registrations and after registrations the users can login with their authentication.

Where after the existing users can send messages to privately and publicly, options are built. Users can also share post with others. The user can able to search the other user profiles and public posts. In this module users can also accept and send friend requests.

With all the basic feature of Online Social Networking System modules is build up in the initial module, to prove and evaluate our system features.

High Profile User:

High profile user who is not normal user like (news channel page, sports person. etc) they have all features of normal user.

High profile User can create tweets and this tweet will be displayed to his/her followers.

Before sending each message is verified using Behavior profile, if message don't match with behavior profile message is blocked and alert message is sent to high profile user.

Admin Module:

Admin can view users who are registered and admin can authorize users. Admin can see all friend requests information. Admin can see all tweets and rewets Messages, which come from a user's weekly messages in timeline, form a time series. To model a user as a subject of series of tweets, we apply COMPA which has large learning capacity to detect malicious user.

Behaviour Profile Calculation:

A behavioral profile for a user U is built in the following way: Initially, our system obtains the stream of messages of U from the social networking site. The message stream is a list of all messages that the user has posted on the social network, in chronological order. To be able to build a comprehensive profile, the stream needs to contain a minimum amount of messages. In our experiments, we empirically determined that a stream consisting of less than $S = 9$ messages does usually not contain enough variety to build a representative behavioral profile for the corresponding account.

Our approach models the following three features when building a behavioral profile in COMPA ALGORITHM.

Time (hour of day). This model captures the hour(s) of the day during which an account is typically active. Many users have certain periods during the course of a day where they are more likely to post (e.g., lunch breaks) and others that are typically quiet (e.g., regular sleeping hours). If a user's stream indicates regularities in social network usage, messages that appear during hours that are associated with quiet periods are considered anomalous.

Message Source The source of a message is the name of the application that was used to submit it. Most social networking sites offer traditional web and mobile web access to their users, along with applications for mobile platforms such as iOS and Android. Many social network ecosystems provide access to a multitude of applications created by independent, third-party developers.

Message Topic. Users post many messages that contain chatter or mundane information. But we would also expect that many users have a set of topics that they frequently talk about, such as favorite sports teams, music bands, or TV shows. When users typically

focus on a few topics in their messages and then suddenly post about some different and unrelated subject, this new message should be rated as anomalous.

4. SCREEN SHOTS



Fig 4.1: Admin Login Page



Fig 4.2: Admin Home

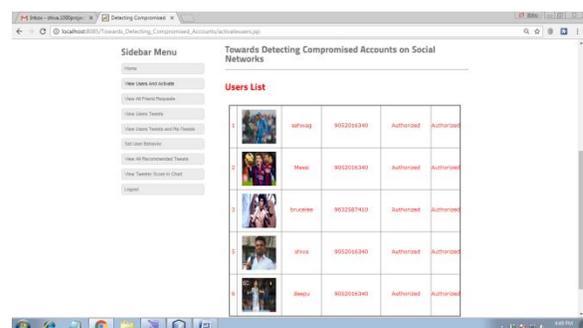


Fig 4.3: View users and activate

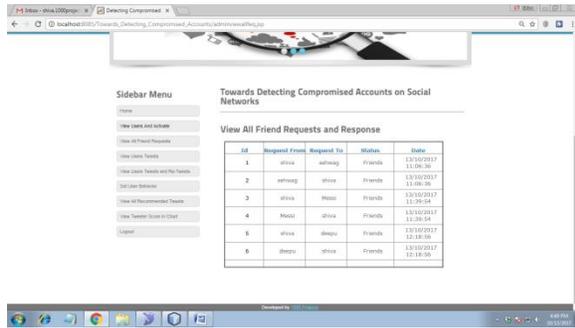


Fig 4.3: View all friend requests

5. CONCLUSION AND FUTURE SCOPE

In this paper, we presented COMPA, a system to detect compromised accounts on social networks. COMPA uses statistical models to characterize the behavior of social network users, and leverages anomaly detection techniques to identify sudden changes in their behavior. The results show that our approach can reliably detect compromises affecting high-profile social network accounts, and can detect compromises of regular accounts, whose behavior is typically more variable, by aggregating together similar malicious messages.

6. REFERENCES

- [1] T. Jagatic, N. Johnson, M. Jakobsson, and T. Jagatif, "Social Phishing," *Comm. ACM*, vol. 50, no. 10, pp. 94–100, 2007.
- [2] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: the underground on 140 characters or less," in *ACM Conference on Computer and Communications Security (CCS)*, 2010.
- [3] "Fox news's hacked twitter feed declares obama dead," <http://www.guardian.co.uk/news/blog/2011/jul/04/fox-news-hacked-twitter-obama-dead>, 2011.
- [4] "U.s. stocks tank briefly in wake of associated press twitter account hack," <http://allthingsd.com/20130423/u-s-stocks-tank-briefly-in-wake-of-associated-press-twitter-account-hack/>.
- [5] "Skype twitter account hacked, anti-microsoft status retweeted more than 8,000 times," <http://www.theverge.com/2014/1/1/5264540/skype-twitter-facebook-blog-accounts-hacked>, 2014.
- [6] E. Lee, "Associated press Twitter account hacked in marketmoving attack," <http://www.bloomberg.com/news/2013-04->



23/ dow-jones-drops-recovers-after-false-report-on-ap-twitter-page. html, 2013.

[7] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, “Detecting Spammers on Twitter,” in Conference on Email and Anti-Spam (CEAS), 2010.