Techniques to Detect Spammers in Twitter- A Survey

Monika Verma
Ph.D. Scholar
Department of Computer Science
PEC University of Technology
Chandigarh, India

Divya, Ph.D
Associate Professor
Department of Computer Science
PEC University of Technology,
Chandigarh, India

Sanjeev Sofat, Ph.D
Professor
Department of Computer Science
PEC University of Technology,
Chandigarh, India

ABSTRACT

With the rapid growth of social networking sites for communicating, sharing, storing and managing significant information, it is attracting cybercriminals who misuse the Web to exploit vulnerabilities for their illicit benefits. Forged online accounts crack up every day. Impersonators, phishers, scammers and spammers crop up all the time in Online Social Networks (OSNs), and are harder to identify. Spammers are the users who send unsolicited messages to a large audience with the intention of advertising some product or to lure victims to click on malicious links or infecting user's system just for the purpose of making money. A lot of research has been done to detect spam profiles in OSNs. In this paper we have reviewed the existing techniques for detecting spam users in Twitter social network. Features for the detection of spammers could be user based or content based or both. Current study provides an overview of the methods, features used, detection rate and their limitations (if any) for detecting spam profiles mainly in Twitter.

Categories and Subject Descriptors

[General Literature]: Introductory and Survey
[Social Networks]: Security

General Terms

User based features, Content based features, Accuracy, Spam profiles, Malicious users.

Keywords

Online Social Networks (OSNs), Twitter, Spammers, Legitimate users.

1. INTRODUCTION

According to Boyd et al. [5] a social networking site allows its users to (a) construct a profile (b) befriend with a list of other users (c) analyze and traverse own and other's list of friends. These Online Social Networks (OSNs) use Web 2.0 technology, which allows users to interact with each other. These social networking sites are growing rapidly and changing the way people keep in contacts with each other. In less than 8 years, these sites have shifted from a forte of online activity to a phenomenon in which millions of internet users are engaged. Online communities bring people with same interests together which makes them easier to keep in contacts with others easily.

Social networking sites [5] started with sixdegrees.com in 1997 and then came up makeoutclub.com in 2000. Sixdegrees.com and other such sites couldn't survive much and disappeared very soon but new sites like MySpace, LinkedIn, Bebo, Orkut, Twitter etc. became successful. Facebook-the very famous site was launched in 2004 [5] and gained a lot of popularity in the world. With larger user databases in OSNs, they are becoming more interesting targets for spammers/malicious users. Spam can take different forms on social web sites and is not easy to be detected. Anyone who is familiar with Internet has faced spam of some sort, be it e-mail spam, spam on forums, newsgroups etc. Spam [18] is defined as the use of electronic messaging system to send unsolicited bulk messages. With the rise of OSNs, it has become a platform for spreading spam. Spammers intend to post advertisements of products to unrelated users. Some spammers post URLs as phishing websites which are used to steal user's sensitive data.

Many papers have been published on the detection of spam profiles in OSNs. But so far no review paper has been published in this field which consolidated the existing research. Our paper aims to provide a review of the academic research and work done in this field by various researchers and highlight the future research direction. In this paper the techniques available for detection of spammers in Twitter have been presented along with their analysis and comparison. This paper is structured as follows: Section 2 describes methodology used to carry out this review; followed security issues in OSNs which have been briefed in Section 3; Section 4 presents definition of spammers and their motives; Introduction to Twitter and its threats has been covered in Section 5; Section 6 is about the motivation behind this survey paper; Section 7 covers the attributes that can be used for detection purpose; Section 8 reviews the work done by various researchers with a comparative analysis; Section 9 gives research directions for new researchers; finally Section 10 concludes the review.

2. METHODOLOGY

This survey of existing methods for detecting spam profiles in OSNs has been done after a systematic review with principled approach in which major research databases for Computer Science have been searched like IEEE Xplore, ACM Digital Library, SpringerLink, Google Scholar, ScienceDirect for concerned topic. We focussed on papers after year 2009 only as the concept of social networks came into existence only in 1997 [1] and became popular only later. Then Facebook was launched in the year 2004 [1] which became very popular. So it took some time for people to get familiar with these networks for communication and hence the attacks on these networks.

This search from above mentioned 5 major databases returned over 60 papers. Papers reviewed for this survey paper were selected after reading titles and abstracts of all the papers. Only those papers were chosen that were found suitable for the present study. Papers with titles and abstracts regarding spam messages detection and other irrelevant topics are excluded for the present paper so finally a total of 21 papers have been selected for review. Mainly the papers have been categorized on the basis of features used to detect spammers.

Through this paper we are trying to compile a list of social networking papers on detection of spam profiles in Twitter that we have read. The list may likely be incomplete, but gives shape to the current research surrounding social network spammer detection. After going through this survey paper, new researchers can easily evaluate what work has been done, in which year and how the present work can be extended to make spam detection more accurate. Whenever appropriate, we have detailed the methodology followed; dataset used; features for detection of spammers and accuracy of the techniques being used by various authors.

In particular, the papers cover how spammers engage with social network users, their implications and existing techniques to detect these spammers.

3. SECURITY ISSUES IN OSNs

Online Social Networking sites (OSNs) are vulnerable to security and privacy issues because of the amount of user information being processed by these sites each day. Users of social networking sites are exposed to various attacks:

1) Viruses – spammers use the social networks as a platform [19] to spread malicious data in the system of users.
2) Phishing attacks - user's sensitive information is acquired by impersonating a trustworthy third party [30].
3) Spammers - send spam messages to the users of social networks [11].
4) Sybil (fake) attack - attacker obtains multiple fake identities and pretends to be genuine in the system in order to harm the reputation of honest users in the network [20].
5) Social bots- a collection of fake profiles which are created to gather users' personal data [32].
6) Clone and identity theft attacks- where attackers create a profile of already existing user in the same network or across different networks in order to fool the cloned user's friends [23]. If victims accept the friend requests sent by these cloned identities, then attackers will be able to access their information. These attacks consume extra resources from users and systems.

4. TYPES OF SPAMMERS

Spammers are the malicious users who contaminate the information presented by legitimate users and in turn pose a risk to the security and privacy of social networks. Spammers belong to one of the following categories [22]:

1. Phishers: are the users who behave like a normal user to acquire personal data of other genuine users.
2. Fake Users: are the users who impersonate the profiles of genuine users to send spam content to the friends' of that user or other users in the network.
3. Promoters: are the ones who send malicious links of advertisements or other promotional links to others so as to obtain their personal information.

Motives of Spammers:

a) Disseminate pornography
b) Spread viruses
c) Phishing attacks
d) Compromise system reputation

5. TWITTER AS AN OSN

5.1 Introduction

Twitter is a social network service launched in March 21, 2006 [14] and has 500 million active users [14] till date who share information. Twitter uses a chirping bird as its logo and hence the name Twitter. Users can access it to exchange frequent information called 'tweets' which are messages of up to 140 characters long that anyone can send or read. These tweets are public by default and visible to all those who are following the tweeter. Users share these tweets which may contain news, opinions, photos, videos, links, and messages. Following is the standard terminology used in Twitter and relevant to our work:

- **Tweets [3]:** A message on Twitter containing maximum length of 140 characters.
- **Followers & Followings [3]:** Followers are the users who are following a particular user and followings are the users whom user follows.
- **Retweet [3]:** A tweet that has been reshared with all followers of a user.
- **Hashtag [3]:** The # symbol is used to tag keywords or topics in a tweet to make it easily identifiable for search purposes.
- **Mention [3]:** Tweets can include replies and mentions of other users by preceding their usernames with @ sign.
- **Lists [3]:** Twitter provides a mechanism to list users you follow into groups
- **Direct Message [3]:** Also called a DM, this represents Twitter's direct messaging system for private communication amongst users.

As per Twitter policy [16], indicators of spam profiles are the metrics such as following a large number of users in a short period of time[1]or if post consists mainly of links or if popular hashtags (#) are used when posting unrelated information or repeatedly posting other user's tweets as your own. There is a provision for users to report spam profiles to Twitter by posting a tweet to @spam. But in Twitter policy [16] there is no clear indication of whether there are automated processes that look for these conditions or whether the administrators rely on user reporting, although it is believed that a combination approach is used.

5.2 Threats on Twitter

1. **Spammed Tweets [13]:** Twitter allows its users to post tweets of maximum 140 characters but regardless of the character limit, cybercriminals have found a way to actually use this limitation to their advantage by creating short but compelling tweets with links for promotions for free vouchers or job advertisement posts or other promotions.
2. **Malware downloads [13]:** Twitter has been used by cyber criminals to spread posts with links to malware download pages. FAKEAV and backdoor[13] applications are the examples of Twitter worm that sent
direct messages, and even malware that affected both Windows and Mac operating systems. The most tarnished social media malware is KOOBFACE [13], which targeted both Twitter and Facebook.
3. **Twitter bots [13]:** Cybercriminals tend to use Twitter to manage and control botnets. These botnets control the users' accounts and pose a threat to their security and privacy.

6. Social Implications of OSNs

Along with the usual problems like spamming, phishing attacks, malware infections, social bots, viruses etc., the greater challenge

that social networking sites present for users is to keep private data secure and confidential.

[1] According to Twitter policy [17], if the number of followings of an account is exceeding 2,000, this number is limited by the number of the account's followers.

The purpose of social networking sites is to make information easily available and accessible to others. But regrettably, cyber criminals use this publicly available information to carry out targeted attacks. Once attackers get access to one of user's accounts, they can easily find a way to excavate more information and to use this information to access their other accounts and accounts of their friends.

6. MOTIVATION BEHIND REVIEW

Because of the ease of sharing information and to be in sync with ongoing topics, Social Networks have become a target for spammers. Detecting such malicious users in OSNs is difficult as spammers are very well aware of the techniques available to detect them. OSNs provide a perfect platform for spammers to disguise as a genuine user and try to get malicious posts clicked by normal users for sake of making money. So detecting such users in order to make network secure and keep the private information of users confidential is the most important topic being delved into by various researchers. So this paper will be very helpful for researchers to swiftly review the work that has been done in this area.

7. FEATURES DISTINGUISHING SPAMMERS & NON-SPAMMERS IN TWITTER

Table 1 lists the publications reviewed in this paper and the category of features used for detection of spam profiles in Twitter. Features on the basis of which spam and non spam profiles are differentiated are user based or content based. User based features are the properties of the profile and the behaviour of user in any social network and content based features are the properties of the text posted by users.

Table 1. Features for the detection of spam profiles

Attributes used for detection of spam profiles
User based features: Which include demographic features like profile details, number of followers, number of followings, followers/following ratio, reputation, age of account, avg. time between tweets posting time behaviour, idle hours, tweet frequency etc.[33,12,34,3,26]
Content based features: Whic include number of hashtags(#), number of URLs in tweets, @ mentions, retweets, spam words, HTTP links, trending topics, duplicate tweets etc.[33,7,11,25]
User based and content based both [1,22,24,27,29,2,4]
Any other feature like graphical distance, graph connectivity: Markov clustering method, URL rate, interaction rate, social relations, social activities, graph based features, neighbor based features, automation based features [21,9,28,33,23,6]

Role of above mentioned features for spam profile detection as per Twitter policy [16]:

1. Numbers of followers-spammers have less number of followers.
2. Numbers of followings-Spammers tend to follow a large number of users.
3. Followers/Following Ratio- this ratio is less than 1 for spammers.
4. Reputation is defined as the ratio of followers to the sum of followers and followings. Spammers have reputation<1.
5. Age of account- is obtained from current date and account creation date. Spammers have generally new accounts so this feature has less value for spammers.
6. Avg. time between posts- spammers post more tweets in a short period of time in order to gain other's attention.
7. Posting time behaviour- spammers tend to post at fixed time schedule may be early morning or late night when genuine users don't use SNS.
8. Idle hours- spammers keep sending messages so they have less idle hours.
9. Tweet frequency- spammers post tweets more frequently at odd times to get attention of other users.
10. No. of hashtages(#)- spammers tweet multiple unrelated updates to the most mentioned topics on Twitter using # to lure legitimate users to read their tweets.
11. No. of URLs- spammer's tweets consist of large number of URLs of malicious sites.
12. @mentions- spammers use maximum @usernames of unknown users in their tweets so as to avoid being detected.
13. Retweets- Retweets are the replies to any tweet using @RT symbol and spammers use maximum @RT in their tweets.
14. Spam Words- Spammer's tweets mainly consist of spam words.
15. HTTP links- if tweets contain maximum number of www or http://, then they are posted by spammers.
16. Duplicate tweets- spammers tend to post duplicate tweets with different @usernames in tweets.

8. EXISTING METHODS FOR DETECTION OF SPAM PROFILES IN TWITTER

Different techniques have been used by researchers to find out the spam profiles in various OSNs. We are focussing only on the work that has been done to identify spammers in Twitter as it is not only a social communication media but in fact is used to share and spread information related to trending topics in real time. Table 2 is showing the summary of the papers reviewed regarding the detection of spammers in Twitter.

Table 2. Outline of techniques used for the detection of spammers

Author	Metrics Used	Methodology Used	Dataset Used	Results
Alex Hai Wang[1]	Graph Based and Content based	Compared Naive Bayesian, Neural Network, SVM and Decision Tree	Validated on 500 Twitter users with 20 recent tweets	Naive Bayesian giving highest accuracy - 93.5%
Lee et. al.[22]	User based	Compared Decorate, SimpleLogistic, FT, LogiBoost, RandomSubSpace, Bagging, J48, LibSVM	Validated on 1000 Twitter users	Decorate giving highest accuracy- 88.98%
Benevenuto et.	User based	SVM	Validated on 1065	Accuracy- 87.6% (with

al.[7]	and Content based		Twitter users	user based and content based features) and accuracy-84.5% (with only user based features)
Gee et. al.[12]	User based	Compared Naive Bayesian, SVM	Validated on 450 Twitter users with 200 recent tweets	Accuracy-89.6%
McCord et. al.[24]	User based and content based	Compared Random Forest, SVM, Naive Bayesian, K-NN	Validated on 1000 Twitter users with 100 recent tweets	Radom Forest giving highest accuracy-95.7%
Lin et. al.[28]	URL rate, interaction rate	J48	Validated on 400 Twitter users	Precision-86%
Amit A. et. al.[2]	Introduced 15 new features	Compared Random Forest, Decision Tree, Decorate, Naive Bayesian	Validated on 31,808 Twitter users	Accuracy-93.6%
Chakraborty et. al.[4]	User based, Content based	Compared Random Forest, SVM, Naive Bayesian, Decision Tree	Trained on 5000 Twitter users with 200 recent tweets	SVM giving highest accuracy-89%
Yang et. al.[6]	18 features (8-existing & 10 new features introduced)	Compared Random Forest, Decision Tree, Decorate, Naive Bayesian	Validated on two datasets-5000 users and then 3500 users with 40 recent tweets	Bayesian giving highest accuracy-88.6%

Significant work has been done by Alex Hai Wang [1] in the year 2010 which used user based as well as content based features for detection of spam profiles. A spam detection prototype system has been proposed to identify suspicious users in Twitter. A directed social graph model has been proposed to explore the "follower" and "friend" relationships. Based on Twitter's spam policy, content-based features and user-based features have been used to facilitate spam detection with Bayesian classification algorithm. Classic evaluation metrics have been used to compare the performance of various traditional classification methods like Decision Tree, Support Vector Machine (SVM), Naive Bayesian, and Neural Networks and amongst all Bayesian classifier has been judged the best in terms of performance. Over the crawled dataset of 2,000 users and test dataset of 500 users, system achieved an accuracy of 93.5% and 89% precision. Limitation of this approach is that is has been tested on very less dataset of 500 users by considering their 20 recent tweets.

Lee et. al.[22] deployed social honeypots consisting of genuine profiles that detected suspicious users and its bot collected evidence of the spam by crawling the profile of the user sending the unwanted friend requests and hyperlinks in MySpace and Twitter. Features of profiles like their posting behaviour, content and friend information to develop a machine learning classifier have been used for identifying spammers. After analysis profiles of users who sent unsolicited friend requests to these social honeypots in MySpace and Twitter have been collected. LIBSVM classifier has been used for identification of spammers. One good point in the approach is that it has been validated on two different combinations of dataset – once with 10% spammers+90% non-spammers and again with 10% non-spammers+90% spammers. Limitation of the approach is that less dataset has been used for validation.

Benevenuto et. al. [7] detected spammers on the basis of tweet content and user based features. Tweet content attributes used are - number of hashtags per number of words in each tweet, number of URLs per word, number of words of each tweet, number of characters of each tweet, number of URLs in each tweet, number of hashtags in each tweet, number of numeric characters that appear in the text, number of users mentioned in each tweet, number of times the tweet has been retweeted. Fraction of tweets containing URLs, fraction of tweets that contains spam words, and average number of words that are hashtags on the tweets are the characteristics that differentiate spammers from non spammers. Dataset of 54 million users on Twitter has been crawled with 1065 users manually labelled as spammers and non-spammers. A supervised machine learning scheme i.e. SVM classifier has been used to distinguish between spammers and non spammers. Detection accuracy of the system is 87.6% with only 3.6% non-spammers misclassified.

Twitter facilitates its users to report spam users to them by sending a message to "@spam". So Gee et. al. [12] utilized this feature and detected spam profiles using classification technique. Normal user profiles have been collected using Twitter API and spam profiles have been collected from "@spam" in Twitter. Collected data was represented in JSON then it was presented in matrix form using CSV format. Matrix has users as rows and features as columns. Then CSV files were trained using Naive Bayes algorithm with 27% error rate then SVM algorithm has been used with error rate of 10%. Spam profiles detection accuracy is 89.3%. Limitation of this approach is that not very technical features have been used for detection and precision is also less i.e. 89.3% so it has been suggested that aggressive deployment of any system should be done only if precision is more than 99%.

McCord et.al. [24] used user based features like number of friends, number of followers and content based features like number of URLs, replies/mentions, retweets, hashtags of collected database. Classifiers namely Random Forest, Support Vector Machine (SVM), Naive Bayesian and K-Nearest Neighbour have been used to identify spam profiles in Twitter. Method has been validated on 1000 users with 95.7% precision and 95.7% accuracy using the Random Forest classifier and this classifier gives the best results followed by the SMO, Naive Bayesian and K-NN classifiers. Limitation of this approach is that for considered dataset reputation feature has been showing wrong results i.e. it is not able to differentiate spammers and non-spammers, unbalanced dataset has been used so Random Forest is giving best results as this classifier is generally used in case of unbalanced dataset, and finally the approach has been validated on less dataset.

Lin et. al. [28] detected long-surviving spam accounts in Twitter on the basis of two different features that are URL rate and interaction rate. Most of the papers have used lot many features for detection of spam accounts like no of followers, no of following, followers/following ratio, tweet content, no of hashtags, URL links etc. But as per this paper all these features are not so effective in detecting spammers so only simple yet effective features like URL rate and interaction rate have been used for detection purpose. URL rate is the number of tweets with URL / total number of tweets and interaction rate is the number of tweets interacting / total number of tweets. 26,758 accounts have been crawled using Twitter API and 816 long surviving accounts have been analysed J48 classifier with 86% precision. Limitation of the approach is that only two features have been used for spam profile detection and if spammers keep low URL rate and low interaction rate then this technique will not work as intended.

According to Amit A. et. al. [2] there are two types of spammer detection techniques – users centric which are based on the features related to user like followers/following ratio and another is URL centric which depends on detecting malicious URLs. Approach mentioned in this paper is hybrid which considers above mentioned both types of features. 15 new features have been proposed to detect spammers, along with an alert system to detect spam tweets. Tweet campaigns and techniques used by spammers have also been studied. Two datasets from Twitter have been used one with 500K users and another with 110,789 users. New features that have been used are: Bait oriented features which identify the techniques used by spammers to lure victims to click on malicious links like no of mentions, mentions to non-followers, hijacking trends, intersection with famous trends. Behavioral features include variance in tweet interval, variance in no of tweets per unit time, ratio of variance in tweet interval to variance in no of tweets per unit time, and tweeting sources. URL features include duplicate URLs, duplicate domain names, IP/domain ratio. Content entropy features include dissimilarity of tweet content, similarity between tweets, URL and tweet similarity. Profile features include follower/following ratio, profile's description language dissimilarity. Thereafter all these features have been collected from malicious users as well as benign users which were then given to four supervised learning algorithms like Decision Tree, Random Forest, Bayes Network and Decorate using Weka tool. 93.6% of spammers with false positive rate of 1.8% have been detected with Decorate classifier giving best results. This technique has been shown to outperform Twitter's spammer detection policy. But this technique has been tested on only 31,808 users whereas Twitter is considering millions of users.

Chakraborty et. al. [4] have proposed a system to detect abusive users who post abusive contents, including harmful URLs, porn URLs, and phishing links and divert away regular users and harm the privacy of social networks. Two steps in the algorithm have been used- first is to check the profile of a user sending friend request to other user as for abusive content and second is to check the similarity of two profiles. After these two steps it is supposed to recommend whether the user should accept friend request or not. This has been tested on Twitter dataset of 5000 users which was collected with REST API. Features considered for differentiating abusive and non-abusive users are- profile based, content based and timing based. Classifiers like SVM, Decision Tree, Random Forest and Naïve Bayesian have been used. SVM outperforms all classifiers and model is performing with an accuracy of 89%.

Yang et. al. [6] utilized new features for the detection of spammers in Twitter. Various techniques used by spammers for evasion have been discussed. 10 new detection features including three graph-based features, three neighbor-based features, three automation-based features and one timing-based feature have been proposed as these features are difficult as well as expensive to dodge as they are based on the methods which spammers don't use in order to not being detected and requires more money, resources and time for evasion. A total of 18 features (8 existing and 10 newly introduced) have been used for detecting purpose and these have been tested using classifiers like Random Forest, Decision Tree, Decorate and Bayesian Network. Bayesian classifier performs best with an accuracy of 88.6%. Limitation of this approach is that very less data has been crawled and only a particular type of spammers are being detected with less detection rate which is the lower bound of the spammers present in the dataset.

9. RESEARCH DIRECTIONS

During survey it became quite apparent that a lot of work has been done for detecting spam profiles in different OSNs. Still improvements can be made to get better detection rate by using a different technique and covering more and robust features as deciding parameter. So following are the few conclusions drawn from survey:

1. Since Twitter has millions of active users and this number is constantly increasing. And almost all the authors have used very small testing dataset to see the performance of their approach. So there is a need to increase the testing dataset to see the performance of any approach.
2. Need to develop a multivariate model.
3. Need to develop a method that can detect all kinds of spammers.
4. Need to test the approaches on different combinations of spammers and non-spammers.

10. CONCLUSION

Many methods have been developed and used by various researchers to find out spammers in different social networks. From the papers reviewed it can be concluded that most of the work has been done using classification approaches like SVM, Decision Tree, Naive Bayesian, and Random Forest. Detection has been done on the basis of user based features or content based features or a combination of both. Few authors also introduced new features for detection. All the approaches have been validated on very small dataset and have not been even tested with different combinations of spammers and non-spammers. Combination of features for detection of spammers has shown better performance in terms of accuracy, precision, recall etc. as compared to using only user based or content based features.

11. REFERENCES

[1] Alex Hai Wang, Security and Cryptography (SECRYPT), Don't Follow Me: Spam Detection in Twitter, Proceedings of the 2010 International Conference, Pages 1-10, 26-28 July 2010, IEEE.

[2] Amit A. Amleshwaram, Narasimha Reddy, Sandeep Yadav, Guofei Gu, Chao Yang, CATS: Characterizing Automation of Twitter Spammers, Texas A&M University, 2013, IEEE.

[3] Anshu Malhotra, Luam Totti, Wagner Meira Jr., Ponnurangam Kumaraguru, Virgılio Almeida, Studying User Footprints in Different Online Social Networks ,International Conference on Advances in Social Networks Analysis and Mining, 2012, IEEE/ACM.

[4] Ayon Chakraborty, Jyotirmoy Sundi, Som Satapathy, SPAM: A Framework for Social Profile Abuse Monitoring.

[5] Boyd, Ellison, N. B. (2007), Social network sites: Definition, history, and scholarship, Journal of Computer-Mediated Communication, 13(1), article 11, http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

[6] Chao Yang, Robert Chandler Harkreader, Guofei Gu , Die Free or Live Hard? Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers, RAID'11 Proceedings of the 14th international conference on Recent Advances in Intrusion Detection, Pages 318-337, 2011, Springer-Verlag Berlin, Heidelberg, ACM

[7] Fabricio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida, Detecting Spammers on Twitter, CEAS 2010 Seventh annual Collaboration, Electronic messaging, Anti Abuse and Spam Conference, July 2010, Washington, US.

[8] Fact Sheet 35: Social Networking Privacy: How to be Safe, Secure and Social

[9] Faraz Ahmed, Muhammad Abulaish, SMIEEE, An MCL-Based Approach for Spam Profile Detection in Online Social Networks, IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, 2012.

[10] Georgios Kontaxis, Iasonas Polakis, Sotiris Ioannidis and Evangelos P. Markatos, Detecting Social Network Profile Cloning, 3rd International Workshop on Security and Social Networking, 2011, IEEE.

[11] Gianluca Stringhini, Christopher Kruegel, Giovanni Vigna, Detecting Spammers on Social Networks, University of California, Santa Barbara, Proceedings of the 26th Annual Computer Security Applications Conference, ACSAC '10, Austin, Texas USA, pages 1-9, Dec. 6-10, 2010, ACM.

[12] Grace gee, Hakson Teh, Twitter Spammer Profile Detection, 2010.

[13] http://about-threats.trendmicro.com/us/webattack-Information regarding Twitter threats.

[14] http://en.wikipedia.org/wiki/Twitter-Information of Twitter.

[15] http://expandedramblings.com/index.php/march-2013-by-the-numbers-a-few-amazing-twitter-stats-Regarding statistics of Twitter.

[16] http://help.twitter.com/forums/26257/entries/1831- The Twitter Rules.

[17] http://twittnotes.com/2009/03/, 2000-following-limit-on-twitter.html-The 2000 Following Limit Policy on Twitter.

[18] http://www.spamhaus.org/consumer/definition-Spam Definition.

[19] J. Baltazar, J. Costoya, and R. Flores, “The real face of koobface: Thelargest web 2.0 botnet explained,” Trend Micro Threat Research , 2009.

[20] J. Douceur, “The sybil attack,” Peer-to-peer Systems, pp. 251–260, 2002.[12] D. Irani, M. Balduzzi, D. Balzarotti, E. Kirda, and C. Pu, “Reverse socialengineering attacks in online social networks,” Detection of Intrusionsand Malware, and Vulnerability Assessment , pp. 55–74, 2011.

[21] Jonghyuk Song, Sangho Lee and Jong Kim, Spam Filtering in Twitter using Sender-Receiver Relationship, RAID'11 Proceedings of the 14th International Conference on Recent Advances in Intrusion Detection, vol. 6961, Pages 301-317, 2011, Springer, Heidelberg ACM.

[35]

[22] Kyumin Lee, James Caverlee, Steve Webb, Uncovering Social Spammers: Social Honeypots + Machine Learning, Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval, 2010, Pages 435–442, ACM, New York (2010).

[23] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, Engin Kirda, All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks, International World Wide Web Conference Committee (IW3C2), WWW 2009, April 20–24, 2009, Madrid, Spain, ACM

[24] M. McCord, M. Chuah, Spam Detection on Twitter Using Traditional Classifiers, ATC'11, Banff, Canada, Sept 2-4, 2011, IEEE.

[25] Manuel Egele, Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna, COMPA: Detecting Compromised Accounts on Social Networks.

[26] Marcel Flores, Aleksandar Kuzmanovic, Searching for Spam: Detecting Fraudulent Accounts via Web Search, LNCS 7799, pp. 208–217, 2013. Springer-Verlag Berlin Heidelberg 2013.

[27] Mauro Conti, Radha Poovendran, Marco Secchiero, FakeBook: Detecting Fake Profiles in On-line Social Networks, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012.

[28] Po-Ching Lin, Po-Min Huang, A Study of Effective Features for Detecting Long-surviving Twitter Spam Accounts, Advanced Communication Technology (ICACT), 15th International Conference on 27-30 Jan. 2013, IEEE.

[29] Sangho Lee and Jong Kimz, WARNINGBIRD: Detecting Suspicious URLs in Twitter Stream, 19th Network and Distributed System Security Symposium (NDSS), San Diego, California, USA, February 5-8, 2012.

[30] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer, “Social phishing,” Communications of the ACM , vol. 50, no. 10, pp. 94–100, 2007.

[31] Vijay A. Balasubramaniyan, Arjun Maheswaran, Viswanathan Mahalingam, Mustaque Ahamad, H. Venkateswaran, A Crow or a Blackbird? Using True Social Network and Tweeting Behavior to Detect Malicious Entities in Twitter, 2002, ACM

[32] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, “The socialbotnetwork: when bots socialize for fame and money,” in Proceedings of the 27th Annual Computer Security Applications Conference. ACM,2011, pp. 93–102.

[33] Yin Zhuy, Xiao Wang, Erheng Zhong, Nanthan N. Liuy, He Li, Qiang Yang, Discovering Spammers in Social Networks, Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence.

[34] Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y. Zhao, and Yafei Dai, Uncovering Social Network Sybils in the Wild, Proceedings of the 11th ACM/USENIX Internet Measurement Conference (IMC'11), 2011.