Research Paper

Volume : 4 | Issue : 6 | Special Issue June 2015 • ISSN No 2277 - 8179

Detection of Spammers in Online Social Networks

Engineering

KEYWORDS: Social network, Clustering approach, Honeybots, Community based approach.

Fathima Nabeela Ali — Research Scholar, Department of Computer Science and Engineering , MES College of Engineering, Kerala, India

Shiju Kumar P.S. — Assistant Professor, Department of Computer Science and Engineering, MES College of Engineering, Kerala, India

ABSTRACT

The popularity of Online Social Networks (OSNs) is often faced with challenges of dealing with undesirable users and their malicious activities. One of the common malicious activity is spamming. Spamming is the process of imitating a real user and fraud honest users. The main aim of spamming is to spread viruses, launching phishing attack etc. Hence this paper presents three methods for the detection of spammers. The first method is based on clustering approach which detects the spammers based on the behavior of the users. The detected cluster contains not only spam users but also normal users. The second method is based on honeybots which are used to gain the information from the spammers. The third method is a community based approach which detects spammers based on its community features. When compared the three papers, the third paper performs better than the other two.

I. INTRODUCTION

Online Social Networks (OSNs) are those sites where people can interact with their friends, relatives and colleagues around the globe. OSNs such as Facebook and Twitter have been emerged as a cheap and easily accessible social media. These networking sites allow the users to share their views, ideas and interests. Also, OSNs enable the users to display their profiles, multimedia content and link to other users whom they relate to. So there is a possibility that malicious users may take advantage of the personal details of the legitimate users. Malicious users are highly attracted to these type of networking sites to perform malicious activities. One of the malicious activity in OSNs is spamming where a fake user spreads malwares or viruses to the legitimate users. The main aim of spamming is to promote viral marketing, phishing and scam to harass the trustable users so that their trust in a particular site would decrease.

Another important class of malicious attack in OSNs is social botnets. Botnets are a collection of compromised users which are controlled by a botmaster in a command and control (C&C) form.

These compromised users work together to launch distributed denial of service attack against normal users, phishing, malware distribution, spamming and click fraud.

Since spammers are a great problem makers , it is necessary to identify them and their behavior in OSNs. Once they are detected they can be removed from the network and future spamming can be controlled.

In this paper three spammer detection techniques will be discussed. The first method is clustering based method. In this method a cluster containing both spammer and normal profiles would be detected. This method is based on user interaction and user behavior.

The second method is based on honeybots in which the honeybots again trust of the botmaster to gain information about the botnets.

The third detection method is based on the features of the community formed by the network users. In this method the features of community are extracted to detect the spammers.

II. SPAMMER DETECTION TECHNIQUES

Different techniques have been proposed for spammer detection. Here we are going to discuss about three methods. First method is based on clustering approach, the second on honeybots and the third one on community features.

A. Clustering Based Approach

For each Facebook user his/her profile would be logged. The information which is available as public would be collected. Activities related to friend requests, page likes and links shared are logged. A user is considered as spam based on his visible activities in the network.

In this method the social network is modeled as a weighted graph in which nodes representing user profiles and the edges representing interaction between the profiles. Using the interaction graph the friends who are active, pages liked by the users and the links shared can be estimated [1].

The common active friends of a user is calculated by taking the intersection of his/her friends and the set of friends with whom the user interacted. Honest users usually interacts with small number of friends, that is with the active friends. Whereas the spammers interact with large number of friends and their interaction may be one way.

The common page likes of a pair of users is the intersection of page liked by one and the pages liked by the other. The common pages liked by the normal users would be low as compared to the spammers.

The URLs shared between a pair of users is calculated as a fraction of links commonly shared by them. Based on these three features the weight of an edge is calculated. In order to classify the users an adjacency matrix is used where each cell represents the weight of edge. Markov clustering is applied on this matrix which performs matrix expansion and inflation iteratively to form clusters. The resulting clusters are divided into three categories. The first cluster contains only normal users, the second contain only spammers and the third contains a mix up of both spammer and normal user.

B. Honeybot Based Approach

A system called SODEXO [2] is introduced for mitigating malicious attacks. A honeybot is a fake account on a social network which is capable of impersonating an infected node. This method is a two stage process. The first stage is the deployment of honeybots and the second stage is the exploitation of honeybots.

Deployment of honeybot takes place by creating an account on an OSN. The account has been created to imitate a real user [3]. This account user then sends a set of friend requests to its neighbors. After accepting its request, the honeybot starts to monitor the message traffic of its neighbors. The message can be any personal messages, posts or any links shared. If any of this message is a spam message which has not been blacklisted then the honeybot becomes a mem-

ber of botnet and then starts to proceed to the next stage.

In the second stage of this method, the honeybot tries to gain as much as information from the botnet. The gained information is in the form of command and control messages. In order to obtain the information the honeybots need to gain the trust of the botnet. For gaining trust the honeybot need to respond to the command and control messages. The main objective of honeybots is to minimize harm to legitimate users. This is done by sending the spam messages among themselves. Depending on the sophistication of the botnets, sometimes honeybots can be identified using the techniques discussed in [4].

This method consists of a protection system which provide the policies for security based on the information learned from the second stage.

C. Community Based Approach

The aim of this method is to identify spammers in social networking sites. Like clustering approach, here also the social network is modeled as an interaction graph in which nodes representing profiles and edges representing interaction between the users. The weight of an edge represents the total number of messages sent, links shared, etc. sent from one node to another. The various steps involved in this method are community detection, extracting community features and classification [5].

The first step is the community detection where overlapping communities at the node level are detected using OCTracker algorithm[6]. This algorithm categorize the nodes as important nodes, boundary nodes and outlier nodes which do not belong to a community.

After identifying the communities the next step is to extract their features. Features considered here are community features and topological features which include the role of a node. The role of a node means that whether the node is an important node or a boundary node and the number of communities the node belong to. This method also uses total out-degree and reciprocity node features.

After feature extraction the next step is to classify the nodes as spammers and non spammers. For that a classifier has to be learned from a set of pre-labeled nodes. The features of this labeled nodes form the training data for classifying the nodes. Decision tree and Naive Bayes are used for learning the classifier. After classifying the nodes, the spammer nodes are reported to the system administrator who further decides whether to block them or remove them.

In order to prevent future spamming high level communities are considered. Then above three steps are performed to prevent further spamming.

TABLE I. COMPARISON TABLE

Detection Techniques	Parameters		
	Method Used	**Detectability**	**Reachable to**
Clustering Approach	Clustering	Low	Not all spammers
Honeybot Approach	Honeybots	Low	Not all spammers
Community Approach	Naive Bayes Classifier	Low	Comparatively all spammers

III. PERFORMANCE ANALYSIS

Clustering based approach is based on markov clustering. On applying markov clustering algorithm, the profiles are categorized as spam profiles, normal profiles and a mix up of both. So the detectability is low and this method is not reachable to all spammers.

In honeybot based approach, honeybots are used for detecting botnets. In this method the command and control traffic is difficult to detect. since botnets use normal traffic. So the detectability is low. One of the main limitations of honeybots is their reach, that is not all spammers would target them, and that the honeybots can possibly be deceived if the spammers involve a copy-profile attack.

In community based approach, Naive Bayes classifier is used is used for classifying the spammers. In this method higher level interactions are considered for avoiding future spamming. Hence the detectability is high and this method is reachable to comparatively all spammer.

IV. CONCLUSION

Over the past few years, OSNs have emerged as cheap and popular communication and information sharing media. The popularity of social network is often faced with challenges of dealing with undesirable users and their malicious activity in the social networks. One of the common form of such activity is spamming. Here three methods have been discussed for detecting spammers. Out of three, community based approach performs better than the other two since it avoids future spamming.

ACKNOWLEDGMENT

We thank GOD almighty, who showered His abundant grace on us to make this paper. We express our heartfelt thanks and deep sense of gratitude to our Principal. We extend our deep gratitude to our Head of the Department for her valuable help and support.

REFERENCE

[1] Faraz Ahmed and Muhammad Abulaish, "An MCL-Based Approach for Spam Profile Detection in Online Social Networks," IEEE International Conference on Trust, Security and Privacy in Computing and Communications, June 2012. | [2] Quanyan Zhu, Andrew Clark, Radha Poovendran and Tamer Basar, "Deployment and Exploitation of Deceptive Honeybots in Social Networks," IEEE Conference on Decision and Control, December 2013. | [3] K. Lee, J. Caverlee, and S. Webb, "Uncovering Social Spammers: Social Honeypots + Machine Learning," in Proc. of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 435–442, 2010. | [4] P. Wang, L. Wu, R. Cunningham, and C. C. Zou, "Honeybot Detection in Advanced Botnet Attacks," International Journal of Information and Computer Security, vol. 4, no. 1, pp. 30–51, 2010. | [5] Sajid Yousuf Bhat and Muhammed Abulaish, "Community-Based Features for Identifying Spammers in Online Social Networks," IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, August 2013. | [6] Sajid Yousuf Bhat and Muhammed Abulaish, "Octracker: A Density-Based Framework for Tracking the Evolution of Overlapping Communities in OSNs," IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012.