A Generic Feature-based Detection for Facebook Spamming Groups

Meng-Jia Yen
Taiwan Semiconductor Manufacturing Co. Ltd.
Hsinchu, Taiwan
e-mail: inscy3@hotmail.com

Yang-Ling Hwang
Chung Shan Medical University
Taichung, Taiwan
e-mail: yanling_h@yahoo.com

Cheng-Yu Tsai
Institute for Information Industry
Taipei, Taiwan
e-mail: josephtsai@iii.org.tw

Fu-Hau Hsu
National Central University
Taoyuan, Taiwan
e-mail: hsufh@csie.ncu.edu.tw

Chih-Wen Ou
National Central University
Chunghwa Telecom Co. Ltd.
Taoyuan, Taiwan
e-mail: frankou@cht.com.tw

***Abstract*—Facebook spammers often use Facebook groups to propagate spam because every member will automatically receive a notification of the post when a new message is posted on the group's wall. Hence, a Facebook group which is created to scatter spam is called a *spamming group*. Even though detection of e-mail spam or Web-based spam has been developed for a long period of time, current Facebook mechanisms still cannot efficiently remove spamming groups. In this study, we propose a new spamming group detection approach for Facebook.**

Keywords–Facebook; spamming group; online social network

I. INTRODUCTION

Online social networks (OSNs) provide new platforms for Internet users around the world to communicate with each other. In March 2015, Facebook had 1.44 billion monthly active users [1]. Different from email spamming which can be directly conducted by sending spam to any email addresses, a Facebook user can not directly contact another Facebook user if they are not friends. Even if they are friends, directly sending unwelcome messages to friends can result in message blocking. Hence, Facebook spammers often use Facebook groups instead to propagate spam.

A Facebook group, which is similar to a real world group created for various reasons, is a collection of Facebook users who create a space on Facebook for organizing, sharing information, and exchanging resources for themselves. A Facebook group's *wall* is a Web page of a Facebook group which allows the group members to post text, images, links, or media. Group members can comment and respond directly on these items on the group's wall. By default configuration, when a group member posts on a group's wall, all members belonging to this group will receive a notification automatically.

To be a member of a certain group, a Facebook user can join a group by the following two methods: (1)Go to the desired group and send a request to the administrator(s) of the group. (2)Ask a friend, who has been a member of the desired group, to add him to the group. A user is defined as a *volunteer*, if he is added to a Facebook group through the first method. And a user is defined as an *invitee*, if he is added to a Facebook group through the second method.

A Facebook group member can invite his friends to join his group directly without the invitees' confirmation. Such a convenient invitation mechanism allows a spammer to add compromised user accounts and their friends to a Facebook spamming groups created by the spammer. Then, whenever a new spammer-crafted message is posted on a spamming group, every member receives a notification of the spamming post automatically. Spamming on Facebook significantly differs from the traditional email spam and Web-based spam malware [2]. Significant effort was spent on email spam detection [3] [4] in recent years, but few studies have focused on understanding the spamming activities in Facebook groups. Most previous spam-related studies identify email spam based on pattern/signature filtering strategies or manual user report mechanism [5]. However, according to Rahman et al. [2], there is only a low overlap (10%) between the keywords associated with email spam and those they found on Facebook. Besides, photos are more frequently used in Facebook spam. Because Facebook spam has different properties than e-mail spam, existing email spam detection solutions are not suitable for Facebook spamming group detection. There are few studies discussing about how to prevent spamming on Facebook. Gao et al. [6] detect and characterize spam campaigns by using wall messages on the Facebook. You [7] implemented a text filtering mechanism to classify groups by using specific keywords. Facebook currently provides a report mechanism for users to report spamming groups when they think that some groups have obviously spam contents or any other unwelcome contents. Spamming activities violate Facebook's Community Standards. But a report [8] shows that the current report mechanism of Facebook, which heavily relies on the cooperation of users, is not effective in removing spamming groups. Our experiments also show that many active spamming groups survive at least for five months (between December 2013 and April 2014). As a result, it is an important issue to develop a new approach to detect Facebook spam.

In this study, we propose a new approach to detect spamming groups according to their features. To this end, four of these features are targeted by spamming group detection including relationships among members, and members' social activities in a Facebook group. The rest of this paper is

 ISBN: 978-1-61208-493-0

organized as follows: Section II describes related work in this field. Section III describes what the system design principles are and what kinds of feature are selected by us for the spamming group detection. Section IV shows the effectiveness of the prototype implementation. Section V addresses that more features could be adopted to improve the detection accuracy. Such adoption will be included in the future work. Section VI concludes this paper.

II. Related Work

This section compares our approach with a text message classifier [7], which filters the text feature (e.g., group's name, description and posts) to find the spamming groups. This text message classifier is easy to be bypassed because the groups' name and description can be modified at any time. Moreover, the keywords used in email spam significantly differ from those used on Facebook [2]. This classifier needs a large database, which must be maintained continuously. Our mechanism does not rely on keywords and databases. We only use the training data (about 200 samples) to keep our approach working without extra storage and resources. Therefore, we demonstrated that our mechanism can effectively detect spamming groups.

III. System Design

After observing diverse Facebook spamming groups and surveying various reports, we found that spamming groups have the following special features. These features consider not only the relationships among members of a group (e.g., information provided in the invitation record of a group) but also characteristics of social activities made by members in a group (e.g., number of clicks on the post "like" buttons made by normal users). These features play important roles in identifying a spamming group in our system.

Spamming group owners may use compromised accounts or use social techniques to entice normal users to add their friends [9] to a spamming group. If a spamming group has relatively few members, the impact of its spam will be reduced. The more members a spamming group has, the more impact its spam can produce. Hence, the member number of a group can be an factor indicating the influence of a post of the group. Most spamming groups either do not allow members to post any kind of messages on the groups' walls or require that posts from members must be approved by group administrators before appearing on the walls. Some spamming groups may allow members to post messages. However, the posts may be deleted quickly to keep spam on the top of walls. Compared to literal posts, image posts are easier to catch readers' eyes. In order to achieve a better effect of propaganda, a spammer would like to post an image post rather than a plain text post. The proportion of volunteers to invitees in a spamming group is significantly less than the proportion of volunteers to invitees in a normal group. This finding is intuitive because normal users seldom like to voluntarily join an unwelcome spamming group. Users always prefer to spend time on something that actually attracting them. For example, if someone is not interested in a post, it is unlikely that he will click the like button of the post. Annoying messages posted by spammers usually get very few number of "like" button clicks made by normal users.

Our approach detects a spamming group based on the group features described in Table I. These features include propagation ability, attractiveness, posting permission, and social impression. A *liker* of a group is a member of the group who has clicked the like button of a post on the group wall. Instead of calculating the number of clicks on the like buttons of all posts on a group wall, we calculate the distinct likers of all posts in the group so that even if a user has clicked the like button of every post on a group wall, this user is still counted as one liker.

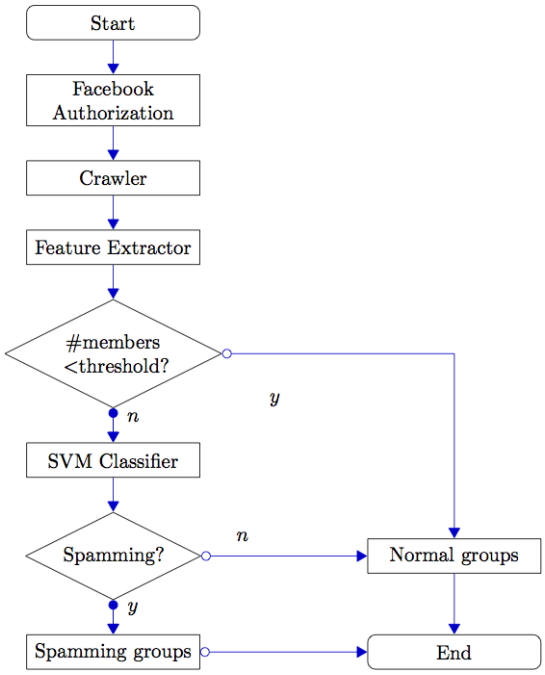

Figure 1. Prototype System Flow chart

The purpose of this study is to develop a prototype system which can identify spamming groups. Figure 1 illustrates the flow chart of our prototype system. Firstly, our prototype system extracts features from a Facebook group specified by a user. After extracting features, our prototype system assesses the number of members of this group. As we have discussed, a typical spamming group is unlikely to have a small number of members. If the number of members is less than a given small threshold, it can be directly classified as a normal group. Even though we might misjudge a spamming group with few members as a normal group in the classification with a small threshold, the number of victims suffering from this false negative is relative small. Secondly, if a group is not classified as a normal group, it is delivered to classifier, which performs classification based on the features listed in Table I. In this prototype system, we use a support vector machine (SVM) as our classifier because it is appropriate for a case which only has a small number of features and the number of output classes is only two. In our case, the outputs are classified-as-normal and classified-as-spamming.

IV. Experiment

Our approach requires a crawler and a classifier. The crawler collects information from Facebook. A machine-learning based classifier is trained by the information collected

 ISBN: 978-1-61208-493-0

TABLE I. FEATURES USED IN OUR APPROACH

Index	Feature	Description
1	Propagation ability	the number of members in a group
2	Attractiveness	the proportion of image posts to the total posts in a group
3	Posting permission	the proportion of distinct posters to all posts in a group
4	Social impression	the proportion of distinct likers to all members in a group

TABLE II. SUMMARY OF DATASET

Group type	training	testing	Total
Normal	100	104	204
Spamming	100	232	332

TABLE III. EXPERIMENT RESULTS

Group type	testing	classified as normal	classified as spamming
Normal	104	98	6
Spamming	232	20	212

from Facebook. It will be able to identify new arriving unclassified Facebook group samples in the testing stage. We implemented a prototype in a host installing Microsoft Windows 7 x64 with Intel(R) core(TM) dual core i5-4430@3.00GHz and 8G RAM. The Average Facebook API response time in normal status is under 200 ms [10]. Our prototype was executed five hundred times to train its classifiers and extract group features. Our prototype checks 100 groups within 20 seconds. Compared with other methods, we provided a real-time and more accurate solution to detect spamming groups.

We qualified 536 Facebook groups shown in Table II, collected during a three-month period from December, 2013 to February, 2014. Each collected Facebook group was manually inspected. Those 100 benign groups and 100 spamming groups were used for training the classifier. The rest of collected groups were used for testing its performance. The experiment result shows the false positive rates, false negative rates, and total error rates in Table III. There are six normal groups misclassified as spamming groups, and 20 spamming groups were erroneously identified as normal groups. Therefore, the false positive rate, false negative rate, and the total error rate of our current approach are 5.77%, 8.62%, and 7.73% respectively.

V. DISCUSSION

We only use four features in this current approach. More features can be adopted in the future work. The invitation record is considered a potential useful feature, since the spamming is a typical abuse of the invitation mechanism of a Facebook group. Attackers invite friends of those compromised accounts, and these benign invitees usually do not actively add their friends to these unwelcome spamming groups. Thus, there is no recursive invitation, which means the number of invitees is restricted naturally. This observation leads to one heuristic: the invitation records can indicate whether a group is abusing the invitation mechanisms. Based on this heuristic, the first feature may be the abuse of invitation, defined as the proportion of invitees to all members in a group. This feature is used to measure whether the invitation mechanism of Facebook is abused in a group. After considering the extent of abuse of invitation, we may also assess the structure of invitation relationships of a group. To this end, some scores may be required for such assessment. We believe that the inclusion of more features will improve the accuracy of the detection. This part of enhancement will be included in the future work.

Our approach makes the following contributions. First, our approach provides an accurate mechanism to identify a spamming group which is better than current Facebook report mechanism. Second, our approach greatly increases spammers' cost to build a spamming group. Third, our approach is flexible to adopt new features to classify spamming groups.

VI. CONCLUSIONS

Facebook groups are abused frequently by spammers. In this study we design and implement a prototype to automatically detect spamming groups. We compare the differences of accuracy between two feature sets. One set contains the four accessed features and the other contains all seven features. Experimental result shows that when only the four accessed features are used, the total error rate of this prototype system is 7.74%.

REFERENCES

[1] Facebook, "Facebook company info," 2016, http://newsroom.fb.com/company-info/.

[2] M. S. Rahman, T.-K. Huang, H. V. Madhyastha, and M. Faloutsos, "Efficient and scalable socware detection in online social networks," in Presented as part of the 21st USENIX Security Symposium (USENIX Security 12). Bellevue, WA: USENIX, 2012, pp. 663–678. [Online]. Available: https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/rahman

[3] C. Kreibich, C. Kanich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage, "Spamcraft: An inside look at spam campaign orchestration," in Proceedings of the 2Nd USENIX Conference on Large-scale Exploits and Emergent Threats: Botnets, Spyware, Worms, and More, ser. LEET'09. Berkeley, CA, USA: USENIX Association, 2009, pp. 4–4. [Online]. Available: http://dl.acm.org/citation.cfm?id=1855676.1855680

[4] Y. Xie, F. Yu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov, "Spamming botnets: Signatures and characteristics," in Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication, ser. SIGCOMM '08. New York, NY, USA: ACM, 2008, pp. 171–182. [Online]. Available: http://doi.acm.org/10.1145/1402958.1402979

[5] Facebook, "How do i deal with spam?" 2016, https://www.facebook.com/help/217854714899185.

[6] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, "Detecting and characterizing social spam campaigns," in Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ser. IMC '10. New York, NY, USA: ACM, 2010, pp. 35–47. [Online]. Available: http://doi.acm.org/10.1145/1879141.1879147

[7] Y.-S. You, "A study on facebook for spamming group detection," Master's thesis, National Tsing Hua University, August 2013.

[8] Facebook, "What is facebook doing to protect me from spam?" 2016, https://www.facebook.com/help/637109102992723.

[9] N. O'Neill, "The rise of scam facebook groups," 2010, http://www.adweek.com/socialtimes/the-rise-of-scam-facebook-groups/312867.

[10] Facebook, "Platform status," 2016, https://developers.facebook.com/status/.

 ISBN: 978-1-61208-493-0