Journal of Computers Vol. 29 No. 4, 2018, pp. 69-81
doi:10.3966/199115992018082904006

Itus: Behavior-based Spamming Group Detection on Facebook

Fu-Hau Hsu[1], Meng-Jia Yan[1], Kai-Wei Chang[1*], Chih-Wen Ou[1], Hung-Min Sun[2]

[1] Department of Computer Sciences and Information Engineering, National Central University,
No. 300, Jhongda Rd., Jhongli, Taoyuan, 32001, Taiwan
hsufh@csie.ncu.edu.tw, inscy3@hotmail.com, popdata520@gmail.com, frankou@cht.com.tw

[2] Department of Computer Sciences National Tsing-Hua University,
No. 101, Section 2, Kuang-Fu Road, Hsinchu, 30013 Taiwan
hmsun@cs.nthu.edu.tw

Received 3 February 2017; Revised 7 August 2017; Accepted 9 August 2017

Abstract. Facebook spammers often use Facebook groups to propagate spam. A Facebook group member can invite his friends to join the group without the invitees' permission. Such a convenient invitation mechanism allows a spammer to add compromised user accounts and their friends to a Facebook group created by the spammer. Then, whenever a new message is posted on the group's wall, every member will receive a notification of the post automatically. This automatic mechanism applies to all group members no matter whether they know or have ever visited this group. As a result, a spammer can easily create a Facebook group to spread spam. 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. Only 14 of 346 spamming groups we monitored were deleted by Facebook in April 2014. Most of the above 346 spamming groups exist for at least five months during our experimental time. Therefore, it becomes an important issue to develop a new solution to identify spamming groups. In this paper, we propose a behavior-based spamming group detection approach for Facebook, called Itus. Itus has an auxiliary crawling Chrome extension to collect and extract features from Facebook groups. These features include relationships between members and their relevant social activities. These features are used for training Itus support vector machine, a machine learning based classifier that can identify a spamming group efficiently. Experimental results shows that the best total detection error rate of Itus is 3.27%.

Keywords: advertisement, behavior-based approach, classifer, Facebook Spamming Groups, social network, SVM

1 Introduction

Online social networks (OSNs) provide new platforms for Internet users around the world to communicate with each other. In March 2015, Facebook has 1.44 billion monthly active users [10]. This large amount of users makes Facebook an attractive target for attackers with various intentions. Spamming is one of common activities launched by attackers on Facebook [9]. Traditional spamming, such as email spamming, distributes lots of spammer-crafted messages to normal users. The spamming on Facebook similarly involves delivery of unsolicited contents or requests to common users. Different from email spamming which can be directly conducted by sending spam to any email addresses, a Facebook user can not directly contact with 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 to propagate spam instead. A Facebook group is a group

[*] Corresponding Author

of Facebook users who can share information. A Facebook group member can add/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 [12]. A great deal of efforts have been made on email spam detection [8, 15] in recent years, but few studies focus 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 [6]. However, according to Rahman et al. [12], there is only 10% of overlap 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. [7] detects and characterizes spam campaigns by using wall messages on the Facebook. You [16] implemented texture 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 [5] shows that the current report mechanism of Facebook, which is heavily relied on the cooperation of users, is not effective to remove 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.

1.1 Background

Before discussing the spamming group, it is necessary to understand some specific terms used by Facebook such as post, like, wall, and group, etc. A post represents the basic unit of information which is often considered as a message and is shared by a poster of Facebook. A post has lots of forms. It can be a pure text message, or a combination of text, images, and even videos. In a Facebook group, a member can leave a literal or an image message as a post on the group's wall only if he has the sufficient privilege to do so. The like button of a post is a special button that allows a Facebook user to express his appreciation. The number of clicks on the like buttons denotes how many Facebook users appreciate it. If a post is attractive to its readers, the post will be very likely to earn lots of "like" clicked by its readers.

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. Besides, a group wall also allows the members of the group to raise questions and to schedule events of the group. Group members can comment and response 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 notification automatically.

To be a member of a certain group, a Facebook user can join a group by the following two methods:

- Go to the desired group and send a request to the administrator(s) of the group.
- Ask a friend, who has been a member of the desired group, to add him to the group.

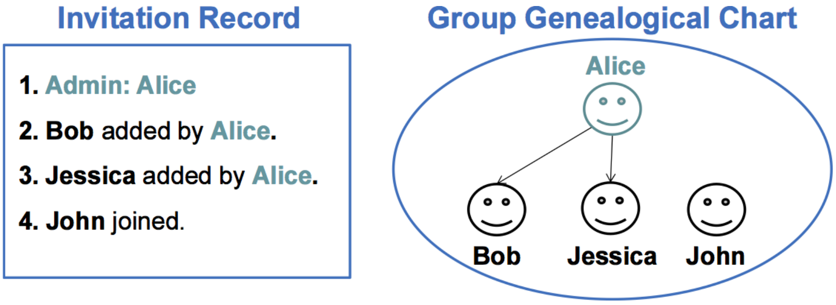
A user is defined as a volunteer, if he is added to a Facebook group through the first method. And a user is defined as an invitee, if he is added to a Facebook group through the second method. Facebook spamming groups result in various problems for Facebook users. First, according to the policy of Facebook, the number of groups that a Facebook user can join is limited. Facebook directly notifies a user of group overuse when he passes the limitation. If the user has already reached this limitation, he needs to leave some joined groups before joining other new ones. As a result, spamming groups decrease the number of benign groups that a user can participate in. Second, a questionnaire analysis [16] shows that the percentage of people who have ever been invited to join a group by their friends is around 98.6%. And 77.8% of users believe that their friends' accounts were compromised when their friends invited them to join a spamming-like group. One-third of users would stop trusting these friends, and even delete these friends from their friend lists. Therefore, spamming groups affect the trust relationships among Facebook users. Third, spamming groups also generate lots of unnecessary Facebook activities and Internet traffic. Fourth, posts from spamming groups are not only annoying, but also possibly damaging.

Using social engineering techniques, many spamming posts are used to sell clothes, electronics, animals, and illegal pharmaceuticals at discounted prices. Finally, some spamming posts try to attract Facebook users either to provide their personal information or to make some transactions without the protection of a trusted online auction system.

1.2 Overview of Itus

In this paper, we propose a new approach, named Itus, to detect spamming groups according to features extracted from group members' behaviors instead of relying on users' reports. Itus is composed of a Facebook API [3] based Web browser extension and a classifier. The browser extension extracts and collects features of a Facebook group specified by its user. The classifier identifies spamming groups based on these features. We use a supervised learning based technique with manually identified normal and spamming samples to train this classifier. These features include static features of a Facebook group, relationships among members, and members' social activities in the group. For example, most Facebook users usually dislike posts made by spammers. Hence, they are unlikely to click the "like" button on a spamming post. As a result, annoying messages posted by a spammer usually get few likes from normal users.

Itus creates the invitation record of a target group to obtain the relationships among the members of a group. The invitation record of a Facebook group describes who the administrator of the group is, who invites others to join the group, who is invited to join the group, and who joins the group voluntarily. For example, Fig. 1 shows the invitation record of a Facebook group. In this group, Alice is the administrator. She invited Bob and Jessica to join this group. But John joined the group voluntarily. Itus uses information provided by the invitation record of a Facebook group to increase the accuracy of Itus. Instead of detecting Sybil accounts directly [13], Itus investigates the invitation record of a group to find out the relationships among members. The genealogical chart of a group is the tree representations of the invitation record of the group. A genealogical chart of a graph consists of several trees. Each node of a tree represents a member of a group. A tree may consist of only one node. If user A invites user B to join a group, there will be an arrow from the node representing user A to the node representing user B. Experimental results show that the invitation records can greatly improve the accuracy of Itus. However, due to privacy concerns, Facebook does not provide the invitation record of a group directly. Thus, we have to obtain such information through crawling.



1. Admin: Alice
2. Bob added by Alice.
3. Jessica added by Alice.
4. John joined.

Fig. 1. Group's genealogical chart of an invitation record

The rest of this paper is organized as follows. Section 2 describes related work in the literature. Section 3 discusses various issues regarding to our system design. Section 4 shows the implementation. Section 5 describes the experimental results of Itus to show the effectiveness and efficiency of Itus. Section 6 discusses possible approach that a spammer may adopt to bypass the detection. Section 7 concludes this paper.

1.3 Contribution

Itus makes the following contributions. First, Itus provides an accurate mechanism to identify spamming group which is better than current Facebook report mechanism. Second, Itus is flexible to adopt new features, and thus greatly increases spammers' cost to build a spamming group against Itus. Itus currently utilizes seven features, it can add more features for further detection without adding too much overhead in the future. If spammers want to bypass Itus with more behavior features, they are very likely to use more fake accounts to achieve that. Such tactics make them more detectable by many existing fake accounts discovering approaches, so that cost of spammers increases dramatically. Third, Itus is a Support Vector Machine approach. In our experiment, we have few features and collect normal amount of training/testing samples. A SVM-based approach can have a reasonable accuracy and runtime performance under these conditions and constraints.

2 Related Work

This section compares Itus with a Facebook advertisement checker [14] and a text message classifier [16]. The Facebook advertisement checker detects spamming groups based on users' reports. It only fetches the basic information of users' groups (i.e., group's ID) and compares the fetched information with its blacklists. The blacklists are established manually by Facebook developers. However, creating spamming groups can be faster than detecting them manually. Some spamming groups in our database are recreated quickly after being deleted by Facebook. The process of updating blacklists is obviously too slow to keep up with the viral propagation of spamming groups on OSNs. The text message classifier [16] filters the text feature (e.g., group's name, description and posts) to find the spamming groups. It 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 [12]. This classifier needs a large database, which must be maintained continuously. Our mechanism does not rely on keywords that a Facebook advertisement checker needs, or a large databases that a text message classifier needs. In our paper, Itus has an auxiliary crawling Chrome extension to collect and extract features from Facebook groups. We only use training data (with about 200 samples) to keep Itus working precisely.

3 System Design

Itus detects Facebook spamming groups based on features frequently occurring in spamming groups. This section introduces the features that we found can be used to detect Facebook spamming groups. This section also describes how we use these features to design our solution and the major components of our solution.

3.1 Features of Spamming Groups

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.

Each spamming group has a large number of members. Spamming group owners may use compromised accounts or use social techniques to entice normal users to add their friends [11] 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 index indicating the influence of a post of the group.

Members' posting permissions are limited by most spamming groups. They prohibit 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. For example, an administrator of a spamming group may allow other members to ask questions about the detail information of what he shared on the wall, but does not want the members to post some entertainment messages, such as sharing of news, funny videos, and photos. The reason why administrators of spamming groups restrict the posting permissions is that messy posts from members will disorder spammers' content. Some spamming groups may allow members to post messages. However, these posts may be deleted quickly by spammers in order to keep their spam on the top of walls.

Posts are usually accompanied with images 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.

Normal users seldom voluntarily join a spamming group. 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.

Only few members actually participate in spamming group activities. Users always prefer to browse something actually attracting them. If a post is not appreciated by the reader, the post is unlikely to obtain a like button from that reader. Annoying messages posted by spammers usually get very few number of "like" button clicks made by normal users.

3.2 Work Flow

The purpose of this study is to develop a prototype system which can identify spamming groups. Fig. 2 illustrates the flow chart of our prototype Itus. First, Itus extracts features from a Facebook group specified by a user. After extracting features, Itus assesses the number of members of this group. According to subsection 3.1, 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. Second, if a group is not classified as a normal group, it is delivered to Itus support vector machine (Itus SVM), which performs classification based on the features discussed in subsection 3.1.

To build Itus SVM, some identified malicious spamming group samples are needed for training this classifier. Itus consists of various modules to handle user authorization, crawling, feature extraction, and classification.

3.3 Itus Components

As shown in Fig. 2, Itus consists of four major modules, user authorization module, crawler module, feature extraction module, and SVM classifier module. A user use to check whether a group he belongs to is a spamming group or not. To make such an examination, the user needs to provide his Facebook account information to Itus firstly, for the authorization module to be authorized by Facebook. Then, the crawler module starts to collect information (such as, group's id, name, and posts on the walls) from the user's groups by invoking Facebook APIs or crawling the walls of the groups of which the user is a member. After the crawling module obtains its information, the feature extraction module begins to extract features of the groups from the crawled information. Finally, the core component of Itus, the SVM classifier module, decides whether a group is a spamming group based on these extracted features.

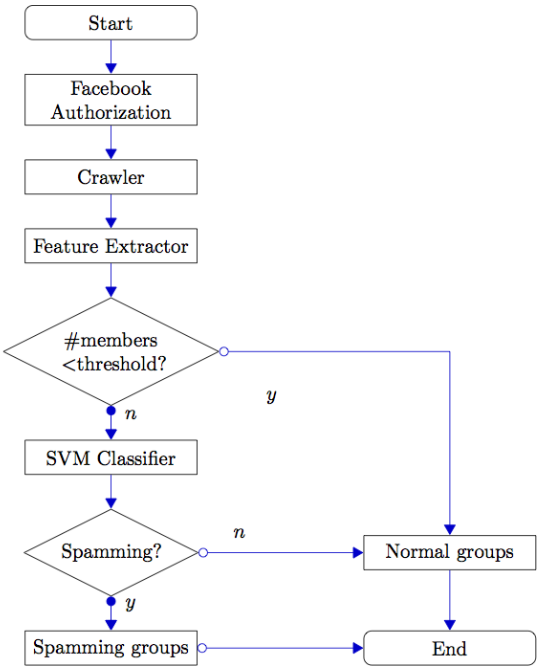


Fig. 2. Flow chart of Itus

3.4 Feature Set

There are two methods to retrieve these features. The first method uses Facebook APIs to retrieve features of a Facebook group. A feature retrieved in this way is called an API-extracted feature. The second method uses Itus crawler module to crawl the wall of a Facebook group to retrieve features of a Facebook group. A feature retrieved in the second way is called a crawler-extracted feature. Table 1 lists a set of API-extracted features. We call this set of features the first feature set and use notation FS_1 to represent this set of features. FS_1 contains four features. Table 2 list a set of both API-extracted features and crawler-extracted features. All API-extracted features in Table 1 also appear in Table 2. We call this set of features the second feature set and use notation FS_2 to represent this set of features. FS_2 contains seven features.

Table 1. Features of FS_1

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

Table 2. Features of FS_2

Feature	Description
Propagation ability	the number of members in a group
Attractiveness	the proportion of image posts to the total posts in a group
Posting permission	the proportion of distinct posters to all posts in a group
Social impression	the proportion of distinct posters to all posts in a group
Abuse of invitation	the proportion of invitees to all members in a group
Member score	accumulated score of all members
Liker Score	accumulated score of all likers

Table 1 includes the following four features, propagation ability, attractiveness, posting permission, and social impression. The first feature, propagation ability, is determined by whether a group has a large number of members. The second one is attractiveness, which is the proportion of the number of the image posts in a group to the number of all posts in the group. The third one is the posting permission derived from the number of distinct posters in a group. A poster of a group is a group member who has made a post on the group wall. The last parameter, social impression, is the proportion of distinct likers to all members in a group. 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 a user has clicked the like button of every post on a group wall, he is still counted as one liker.

Table 2 lists the features of FS_2. FS_2 contains all features of FS_1 and three other different features, abuse of invitation, member score, and liker score; hence, FS_2, is an extension version of FS_1. The crawler module of Itus gathers information from a Facebook group. Then, the feature extraction module retrieves features from the information. Feature “abuse of invitation” is 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. The rest two features, member score and liker score, are used to assess the structure of invitation relationships of a group.

Except the administrator of a Facebook group, there are often two kinds of members in the group. One is the invitee and the other is the volunteer. We use a group genealogical tree to describe the invitation relationships among members of a group. A group genealogical tree may consist of several trees. In a group genealogical chart, a volunteer is represented by the root node of a tree. If a volunteer does not invite other persons to join the group. The tree consists of a single node. And if a member invites another person to become a member of a group, there will be an arrow from the node representing the member to the node representing the invitee. In a tree, the root node is at level 0. Its children are at level one, and so on. A child of a level i node is at level (i+1). Itus focuses on the tree whose root node represents the administrator of a group. We call this tree the group genealogical tree of the group genealogical chart. Fig.

3 is a group genealogical chart. In Fig. 3, member from A to D in level one are all invited by the group administrator, who is the root node of the group genealogical tree. Member from E to H in level two are invited by the members in level one, and so on. In Fig. 3, member from T to Z are volunteers who do not invite other persons to join the group; hence, they are represented by the root nodes of trees that consist of a single node.

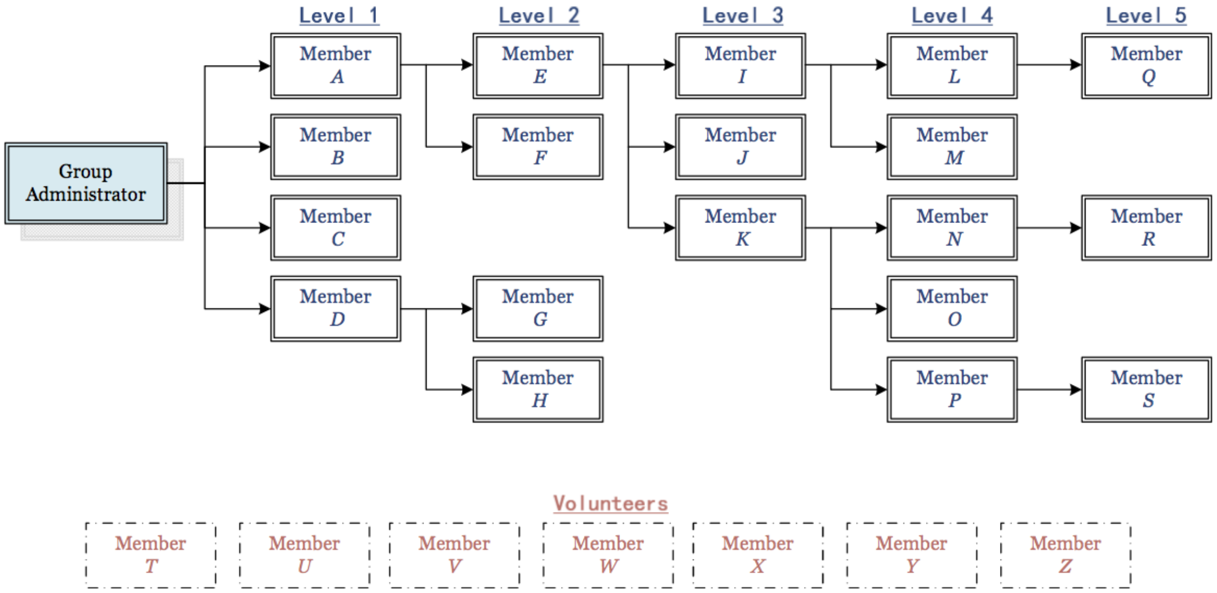


Fig. 3. An example of group genealogical chart

Members in the same level of a group genealogical tree have the same weight. Members in different levels of a group genealogical tree have different weights. All volunteers are deemed as having the same weight as the nodes in the last level of the related group genealogical tree. Itus assigns weight w_i to a member at level i of a group genealogical tree, and weight w_n to a volunteer if the number of levels of the group genealogical tree is n. The member score and liker score of a group are calculated according to the following equations. The member score of a group tries to reflect the fact that a member of a normal group usually will invite his friends to join the group. And in turn, his friends are also very likely to invite their own friends to join the group. As a result, the level of the group genealogical tree of a normal group usually is high. And the probability that a group is a normal group is high, if the group has high member score and/or liker score.

Definition 1. Assume the highest level of a group genealogical tree is n, the weight of each member at level i is w_i, the number of volunteers is m_v, and the number of members at level i is m_i. The member score, $group_{ms}$, of the group is,

$$group_{ms} = \frac{\sum_{i=1}^{n} m_i w_i + m_v w_n}{\sum_{i=1}^{n} m_i + m_v}, 1 \le i \le n. \quad \textbf{(1)}$$

Definition 2. Assume the highest level of a group genealogical tree is n, the weight of each member at level i is w_i, the number of volunteers is m_v, and the number of likers at level i is k_i. The liker score, $group_{ls}$, of the group is,

$$group_{ls} = \frac{\sum_{i=1}^{n} k_i w_i + m_v w_n}{\sum_{i=1}^{n} m_i + m_v}, 1 \le i \le n. \quad \textbf{(2)}$$

To take Fig. 3 as an example, the member score of Fig. 3 is $\frac{(4w_1+4w_2+3w_3+5w_4+3w_5)+7w_5}{26}$.. To calculate the member score, the values of these weights must be determined first. For a group genealogical tree, the number of weights is equivalent to the depth of the group genealogical tree. If the number of wights of a group genealogical tree is n, then the related n weights form a weight vector (w_1, w_2, w_3, …, w_n). We define the weight vector as (1, 2, 3, ..., n) in the member score calculation. The calculation of a liker score uses the same method to assign a value to a weight. After calculating the member score and liker score of a group, all features are ready for the SVM classifier module to use to determine whether a group is a spamming group.

4 Implementation

Itus SVM crawler utilizes Facebook Graph APIs to collect information from Facebook. We represent information by using of nodes (basically "things" such as a group), edges (the connections between those "things" such as a Photo's Comments), and fields (information about those "things" such as the name of a group). Itus obtains the fields shown in Table 4 and Table 5 by making API calls with /group-id, and group-id/feed queries, respectively.

Table 3. Responses of various queries

Query	Response
/user-id/groups	The Facebook groups that a person is a member of.
/group-id	Information of this group, such as id, name, and description.
group-id/feed	The feed of posts (including status updates) and links published on this group.

Table 4. Extracted fields of a group

Property Name	Description	Type
id	The Group ID	string
name	The name of the group	string
description	A brief description of the group	string
owner	The user profile that created this group	user
privacy	The privacy setting of the group	string

Table 5. Extracted fields of a post

Property Name	Description	Type
id	The post ID	string
from	Information about the user profile that posted the message	user
picture	The picture retrieved from any link included in the post	string
like	People who like this post	user

As mentioned in previous subsection, due to privacy concerns, Facebook does not provide APIs to access the following features, abuse of invitation, member score, and liker score; hence, Itus uses its crawler module to extract these features. We developed a Google Chrome extension [2] called an auxiliary crawling program (ACP) to collect features from a group member list document.

After collecting hundreds of normal and spamming group samples, these samples are used to build Itus SVM. Instead of using other popular classifiers, such as decision trees, naive Bayes, and logistic regression, Itus selects SVM to create its classifiers due to the following reasons. First, generic SVM works well for binary classifications, which is equivalent to the normal-spamming group classification of this study. According to Rich et al. [1], SVM is more accurate than the above three classifiers. Second, Itus uses less than ten features. And over five hundred samples are collected in our study. Rich et al. [1] used similar numbers of features and samples as ours (sample case CALHOUS) to evaluate various types of classifiers. According to their bootstrap analysis, SVM ranks fourth among ten classifiers. Third, SVM is easy to implement. SVM has lots of off-the shelf tools for developers. Finally, experimental results show that the total classification error rate of Itus is only 3.27%, which shows that SVM is an accurate tool for our need. However, if it is needed, Itus still could use other classifiers to detect spamming groups.

LibSVM [13] is an efficient tool for SVM classification. Itus uses an SVM extension which wraps LibSVM in a PHP interface for easily using in PHP scripts. Once an SVM has been constructed, it can be used to classify new arriving unclassified samples in the testing stage.

5 Experiment

Itus employs SVMs to identify Facebook spamming groups. Table 6 summarizes data of 550 Facebook groups collected during a three-month period from December, 2013 to February, 2014. In the later stage of data inspection (April, 2014), 14 of 346 spamming groups were deleted by Facebook. Hence, we removed them from our testing data. After checking manually, we found that there were 204 normal Facebook groups and 332 spamming groups in these 536 active Facebook groups. In the training stage, 100 normal groups and 100 spamming groups were used to train the Itus SVM.

Table 6. Summary of dataset

Group type	Number of groups used for training	Number of groups used for testing	Total
Normal	100	104	204
Spamming	100	232	332

In the testing stage, 104 normal groups and 232 spamming groups were used to test the performance and accuracy of Itus. Based on our observation on collected samples, in the testing stage, we chose 200 as the threshold discussed in subsection 3.2. Although there is a threshold to determine whether an inspected group is a spamming group, the number of members of that group is also a feature of Itus SVM.

5.1 Performance

We implemented Itus in a host installing Microsoft Windows 7 x64 with Intel(R) core(TM) i5-4430@3.00GHz CPU and 8G RAM. The Average Facebook API response time in normal status is under 200 ms [4]. Itus was executed five hundred times to train its classifiers and extract group features. The average time for training the classifiers (100 normal groups and 100 spamming groups) was 691ns. The average time for extracting features of a group was 0.186s. Itus could check 100 groups within 20 seconds. Compared with other methods, we provided a real-time and more accurate solution to detect spamming groups.

5.2 Accuracy

As mentioned in Table 1 and Table 2, there are two feature sets, FS_1 and FS_2 of Itus. All features in FS_1 can be extracted by invoking Facebook APIs. FS_2 is an extension version of FS_1. FS_2 includes the features in FS_1 and three features extracted and calculated by the crawler module of Itus. We compared the differences of accuracy between these two feature sets in this subsection. First, we used feature set FS_1 to evaluate the accuracy of Itus. Then, we used feature set FS_2 to evaluate the accuracy of Itus again. In both evaluations, the number of spamming groups used in our testing stage was 232, and the number of normal groups used in our testing stage was 104.

Fig. 4 shows the false positive rates, false negative rates, and total error rates of feature set FS_1 and feature set FS_2. When feature set FS_1 was used by Itus, six normal groups were misclassified as spamming groups, and 20 spamming groups were erroneously identified as normal groups. Therefore, the false positive rate, false negative rate, and the error rate of Itus were 5.77%, 8.62%, and 7.73% respectively. When feature set FS_2 was used by Itus, 4 normal groups were misclassified as spamming groups, and 7 spamming groups were erroneously identified as normal groups. Therefore, the false positive rate, false negative rate, and the error rate of Itus were 3.85%, 3.02%, and 3.27% respectively. The error rate of Itus using feature set FS_2 is less than the error rate of Itus using feature set FS_1. Therefore, the extra features of FS_2 are helpful in increasing the accuracy of Itus.

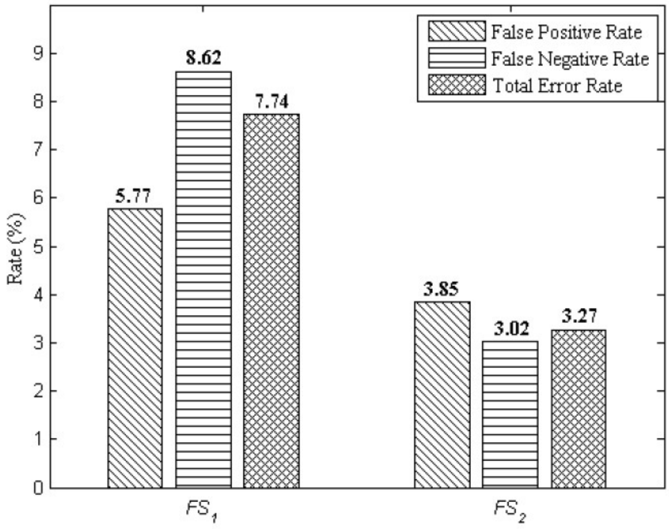


Fig. 4. Results of system evaluation

False Negative. As discussed in previous subsection, when feature set FS_2 was used by Itus, seven spamming groups were misclassified as normal groups. We manually checked these spamming groups to find the reasons that result in the misclassification. Among the seven spamming groups, two spamming groups were created by sellers who had physical stores and had many customers. Hence, many customers either were willing to invite their friends to join the groups, or were willing to add volunteers to the groups. As a result, we think misclassification caused by this reason is not supposed to create problems for Itus users. After all, the related spamming groups are benign and attractive ones. Three of the seven misclassified spamming groups were for open-advertising. A lot of spammers voluntarily to join the groups, and normal members of these group had permissions to publish advertisements. The large number of distinct posters and volunteers resulted in misclassification of Itus. The rest two misclassified spamming groups had few members; hence, Itus deemed them as normal groups. However, the small member number means that the related spamming groups can cause little influence on a small group of users.

False Positive. As discussed in previous subsection, when feature set FS_2 was used by Itus, four normal groups were misclassified as spamming groups. Similarly, we manually checked these normal groups to find the reasons that result in the misclassification. One of these four misclassified normal groups had a small number of posts and the majority of these posts contained images. The rest three misclassified normal groups had a large number of members, but had apparently few social activities. For example, their members seldom clicked the like buttons of the posts on their walls. Such large groups with low social activities are believed to be unusual, and are good cases for further analysis in our future work.

6 Discussion

In this section, we will compare Itus with other existing approaches in 6.1, the limitation of Itus will be discussed in 6.2. Some future work related to enhance Itus will be discussed in 6.3.

6.1 Comparison

In order to compare Itus with other existing similar approaches, we select two papers related to spamming groups detection. We then discuss how they work, and what differences they are in this paragraph. Choo et al., conducted a research whose dataset was collected from Amazon [17]. Those discovered strong positive communities built by review or response activities, were more likely to be opinion spammer groups. Therefore, in that paper, they built general user relationship graphs, representing the users' interaction with one another. Based on review and response activities via sentiment analysis, they can extract abnormally positive relationship graphs to capture boosting behavior and detect spamming groups. The sentiment analysis against the spamming group had been discussed in recent years [20-21]. Such approaches collected a large number of text data from the four popular categories with Amazon's comments. They used the dataset to determine the positive or negative sentiment, and only considered those positive relationships among the user comments as spamming groups. They need to process large amount of content data. Compared to these text-based approaches, Itus has more reasonable runtime performance overhead than these approaches.

The second similar approach [18] analyzed the dataset from Sina Weibo, a Microblogging service provider in China. In this paper, it filtered tweets from January 2010 to June 2011, and detected spamming groups by using of co-retweeting relationships and retweeting content to capture the characteristics of group spammers. It used topic model for user profile construction, and proposed the LDA-G model as benchmarks. If groups were ranked high, they would be considered as spamming groups. The user profile construction of this paper only adopted co-retweeters and retweeting content, which were not difficult for attackers to bypass it. Itus uses seven features, so that attackers have higher cost if they plan to bypass its detection. This paper also discussed some perspectives which might be considered as spammer indicators. Such observation may benefit to strengthen the Itus feature pools as well in the future.

6.2 Limitation

There is a high cost technique that is possible to bypass the detection of Itus. The member score and liker score of a group are two important features that Itus uses to determine whether a group is a spamming group. However, spammers may try to bypass the detection by using more fake member accounts to construct a group genealogical tree with a high depth. As shown in Fig. 5, the administrator of a spamming group can invite fake accounts one by one, to ensure that the last fake account is in a high level. Then, the last fake account invites a huge number of members to join the group, so that the spamming group will not be detected because it has a high member score and liker score. As a result, the two scores of such spamming group are less effective for spamming group identification. We will discuss the solution as our future work in subsection 6.3.

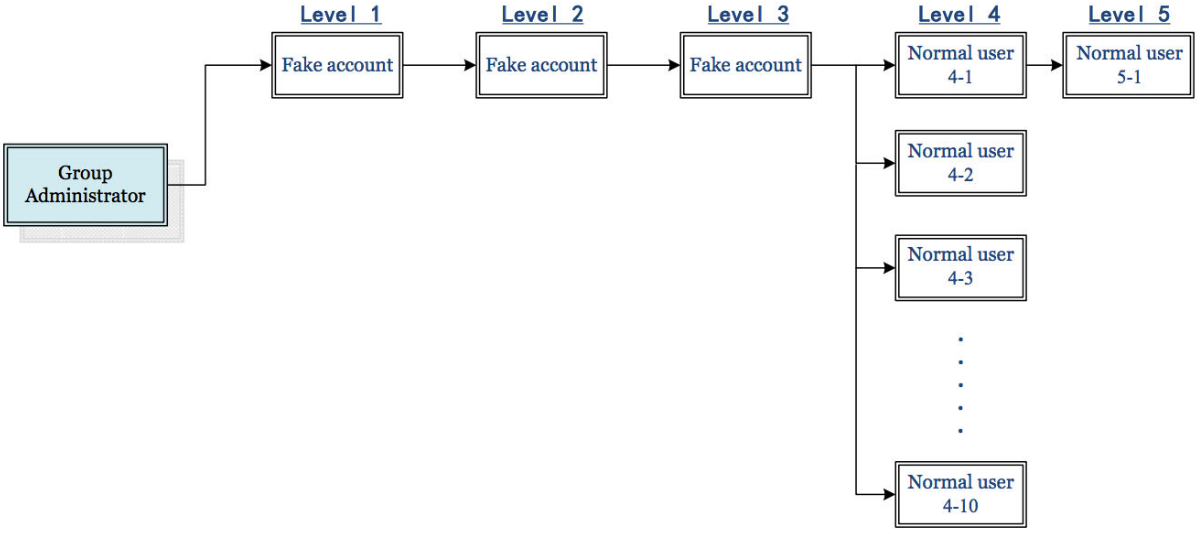


Fig. 5. An example of bypassing Itus detection

Our approach assumes that attacker cannot easily imitate online social behavior pattern of target amount of users. However, for a group of attackers who are well-organized and have sufficient support, they would be able to learn actual social behavior pattern of large amount of users, and simulate the extremely similar social behavior of users to avoid our detection. A study [19] also discussed this limitation. All behavior-based approaches, including Itus, should take it into consideration to explore more features, and to design a more efficient classifier in the future.

6.3 Future Work

In order to solve the problem with fabricated high member score and liker score, Itus can add a new feature: the ratio of the number of internal nodes of a group genealogical tree, to the number of leaf nodes in the tree. The strategical goal of adopting this feature is to force the administrator of a spamming group to use fake accounts as many as possible, if he wants to bypass the detection of Itus. While more fake accounts are used, it has more opportunities to disclose involved spamming groups, so that it certainly increases the spammer's cost and risk. Besides, in our future work, we will work on designing a verification mechanism, similar to these two approaches [22-23], and integrate them with Itus to prevent network behavior from being imitated so easily. Such mechanism is believed to require more features, especially for those features indicating the abuse of fraudulent accounts.

7 Conclusions

Facebook groups are abused frequently by spammers. In this paper we design and implement a prototype, Itus, to automatically detecting spamming groups. Itus is composed of a web browser extension based crawler and Itus SVM. There are four Facebook API accessed features and three extracted features for training Itus SVM. 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 the total error rate of Itus is 7.74% when only the four accessed features are used. The total error rate of Itus decreases to 3.27%, if all seven features are used. Itus has a limiation. For a group of attackers who are well-organized and have sufficient support, they may be able to learn actual social behavior pattern of large amount of users, and simulate the extremely similar social behavior of users to avoid our detection. Such operation takes lots of extra cost for attackers. If the attacker does not have sufficient resources, it can not effectively bypass Itus detection. Such imitation is still a common limitation for all behavior-based approaches. Hence, Itus will explore and use more features, to build a more efficient classifier, and to increase more cost of attackers who plan to bypass Itus in the future.

References

[1] R. Caruana, A. Niculescu-Mizi, An empirical comparison of supervised learning algorithms, in: Proc. the 23rd International Conference on Machine Learning, 2006.

[2] Chrome: Getting started: Building a chrome extension. <https://developer.chrome.com/extensions/getstarted>, 2015.

[3] Facebook: Facebook developers. <https://developers.facebook.com/>, 2014.

[4] Facebook: Platform status. < https://developers.facebook.com/status/>, 2014.

[5] Facebook: What is facebook doing to protect me from spam? <https://www.facebook.com/help/637109102992723>, 2014.

[6] Facebook: How do i deal with spam?, <https://www.facebook.com/help/217854714899185>, 2015.

[7] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, B.Y. Zhao, Detecting and characterizing social spam campaigns. in: Proc. the 10th ACM SIGCOMM Conference on Internet Measurement, 2010.

[8] C. Kreibich, C. Kanich, K. Levchenko, B. Enright, Geoffrey M. Voelker, V. Paxson, S. Savage, Spamcraft: an inside look at spam campaign orchestration, in: Proc. the Second USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET 09), 2009.

[9] W. Luo, J. Liu, J. Liu, C. Fan, An analysis of security in social networks, in: Proc. IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009.

[10] Newsroom., F.: Company info. <http://newsroom.fb.com/company-info/>, 2015.

[11] N. O'Neill, The rise of scam facebook groups. <http://www.adweek.com/socialtimes/the-rise-of-scam-facebook-groups/312867>, 2010 (accessed 13.01.10).

[12] M.-S. Rahman, T.-K. Huang, H.-V. Madhyastha, M. Faloutsos, Efficient and scalable socware detection in online social networks, in: Proc. the 21st USENIX Security Symposium (USENIX Security 12), 2012.

[13] G. Wang, T. Konolige, C. Wilson, X. Wang, H. Zheng, B.Y. Zhao, You are how you click: clickstream analysis for sybil detection, in: Proc. the 22nd USENIX Security Symposium (USENIX Security 13), 2013.

[14] T. Wang, spam-group (2015). <https://github.com/tony1223/spam-group>, 2015.

[15] Y. Xie, F. Yu, K. Achan, R. Panigrahy, G. Hulten, I. Osipkov, Spamming botnets: signatures and characteristics, in: Proc. the ACM SIGCOMM 2008 Conference on Data Communication, 2008.

[16] Y.S. You, A study on Facebook for spamming group detection, [thesis] Hsinchu: National Tsing Hua University, 2013.

[17] E. Choo, T. Yu, M. Chi, Detecting opinion spammer groups through community discovery and sentiment analysis, in: Proc. IFIP Annual Conference on Data and Applications Security and Privacy, 2015.

[18] Q. Zhang, C. Zhang, P. Cai, W. Qian, A. Zhou, Detecting spamming groups in social media based on latent graph, in: Proc. Australasian Database Conference, Databases Theory and Applications, 2015.

[19] K.-S. Adewolea, N.-B. Anuara, A. Kamsina, K.-D. Varathana, S.-A. Razakb, Malicious accounts: dark of the social networks, Journal of Network and Computer Applications 79(2017) 41-67.

[20] J. Wu: Sentiment Analysis. <http://dataology.blogspot.tw/2015/04/sentiment-analysis.html>, 2015.

[21] J. Yadav, A survey on sentiment classification of movie reviews, IJEDR 3(1)(2014) 340-343.

[22] A. Almaatouq, E. Shmueli, M. Nouh, A. Alabdulkareem, V.-K. Singh, M. Alsaleh, A. Alarifi, A. Alfaris, A.-S. Pentland, If it looks like a spammer and behaves like a spammer, it must be a spammer: analysis and detection of microblogging spam accounts, International Journal of Information Security (2016) 1615-5270.

[23] X. Ruan, Z. Wu, H. Wang, S. Jajodia, Profiling online social behaviors for compromised account detection, IEEE Transactions On Information Forensics And Security 11(1)(2016) 176-187.