DESIGN A SYSTEM FOR FIGHTING SOCIAL NETWORKING SPAMMERS

[1]Nivedita Nandgave, [2]Prof. Pooja Khangar

[1]M.Tech. Scholar, [2]Assistant Professor
Electronics Department
Wainganga College of Engineering & Management
Nagpur, India

***Abstract*-Till date, as one of the most popular online social networks (OSNs), Twitter is paying its dues as more and more spammers set their sights on this microblogging site. Twitter spammers can achieve their malicious goals such as sending spam, spreading malware, hosting botnet command and control (C&C) channels, and launching other underground illicit activities. Due to the significance and indispensability of detecting and suspending those spam accounts, many researchers along with the engineers at Twitter Inc. have devoted themselves to keeping Twitter as spam-free online communities.**

Most of the existing studies utilize machine learning techniques to detect Twitter spammers. "While the priest climbs a post, the devil climbs ten." Twitter spammers are evolving to evade existing detection features. We first make a comprehensive and empirical analysis of the evasion tactics utilized by Twitter spammers.

We further design several new detection features to detect more Twitter spammers. In addition, to deeply understand the effectiveness and difficulties of using machine learning features to detect spammers, we analyze the robustness of 24 detection features that are commonly utilized in the literature as well as our proposed ones. Through our experiments, we show that our new designed features are much more effective to be used to detect (even evasive) Twitter spammers.

According to our evaluation, while keeping an even lower false positive rate, the detection rate using our new feature set is also significantly higher than that of existing work. To the best of our knowledge, this work is the first empirical study and evaluation of the effect of evasion tactics utilized by Twitter spammers and is a valuable supplement to this line of research.

I. INTRODUCTION

Online Social Network (OSN) are websites where users can create profiles, establish connection with other users and converse with them. There are hundreds ofsuch OSN websites present today. Facebook, Twitter, etc. are the most popular onesboasting more than 500 million active users. Twitter, as an Online Social Network,is intended to help people converse using text-based posts called tweets. Popularityof Twitter and other OSNs has been rising in recent times having played crucialrole in connecting people and providing a discussion forum on several occasions likeprotests in Syria.

Spammers are users on Twitter which analogous to e-mail spam try to spreadmalicious content or advertise using their social network on Twitter. With riseof Twitter as Online Social Network, it has inevitably attracted large number ofspammers. OSNs have been fighting spammers since their inception. In August2009, Twitter observed around 11% of tweets posted were spam. Twitter has itssocial structure built by users following each other posts which in turn signifies trust
between users. This along with millions of users provide a perfect platform forspammers to disseminate spam.

Spammers not only try to advertise products, they have also been actively involved in deceiving users into clicking malicious links. Spammers on Twitter employmany techniques to lure users into clicking malicious URLs. Techniques whichdeceive users into clicking such URLs include but are not limited to: befriending (tofollow in Twitter terminology) unrelated users and sending unsolicited messages. Togain a wider reach to potential victims, spammers are known to befriend (to followin Twitter terminology) unrelated users, send irrelevent messages and vicious components (for instance, using URL shorteners to substitute malicious appearing URLs.), to convince the victim of their legitimacy. Preventing spam proliferation translates to protecting users from clicking vicious links. The malicious URLs pose threats in the form of drive-by-downloads and other infections. The infected machine may also assist in nefarious botnet activities such as by itself being a source of email spam or used during the execution of Distributed Denial of Service (DDoS) attacks. From Twitter's perspective, spam threatens to prohibit the growth of user base hurting both reputation and revenue. Identifying spammers on Twitter is hard. The problem becomes especially difficult due to resources required to analyze the huge dataset such as that observed by Twitter. An example of such a scale comes from the event where Bin Laden's death spurred Twitter users to generate about 12.4 million tweets an hour. Spammers, in addition, use sophisticated tools which have rendered spam signatures useless. One such tool used by spammers is the Spinbot which generates a sentence with a fixed semantic meaning but varied syntactical structures.

There has been many approaches introduced both in industry and academia tofight spammers on Twitter. Engineers at Twitter Inc. has been working actively to control spammers. They have introduced a set of rules which dictates the behavior of users on Twitter. They define behavior which would be considered as spammer and such accounts are suspended by Twitter.Accordingto Twitter rules, users following/unfollowing aggressively, having lower follower-following ratio, posting duplicate content are considered spammers. However Twitter mentions that definition of spamming behavior will keep on changing depending upon new tricks used by spammers.

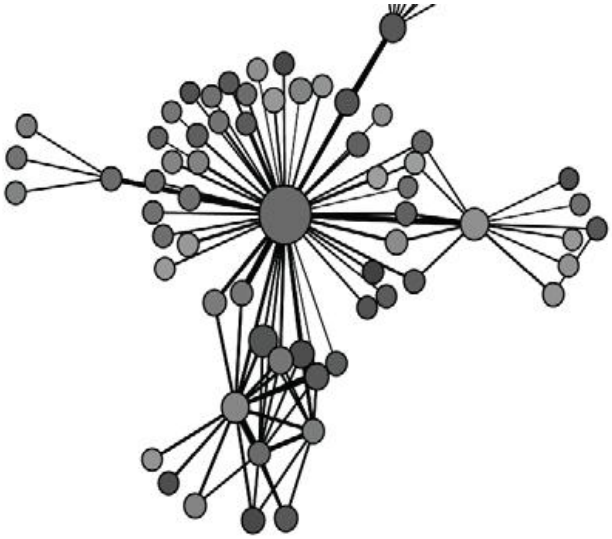

Figure 1.1. A Typical Social Network

II OBJECTIVE

All current Online Social Networks adopt the client-server architecture. The OSN service provider acts as the controlling entity. It stores and manages all the content in the system. On the other hand, the content is generated by users spontaneously from the client side. The OSN service provider offers a rich set of well-defined interfaces through which the users can interact with others. Currently two popular ways of interaction exist. Facebook is representative of OSNs that adopt the interaction between a pair of sender and recipient as their primary way of interaction, although they also support other ways. Twitter is representative of OSNs that adopt broadcasting as their primary way of interaction.

Figure 1(a) illustrates a simplified OSN architecture. It only depicts the components that are related to message exchanging, while all other functionalities,
e.g., user authentication, video sharing and 3rd party applications, are omitted. In this simplified example, multiple users are interacting via the message posting and viewing interface. In Facebook-like OSNs, it represents the case that user A and B are posting messages to user C and D, respectively. In Twitter-like OSNs, it represents the case that user A and B are broadcasting messages to all the followers including user C and D, respectively, while other possible recipients are omitted for simplicity. In both cases, the service provider mediates all the interactions. The generated messages are first stored at the service provider's side, and will be relayed when the corresponding recipient signs in.

Unfortunately, all the content in OSNs is generated by users and is not necessarily legitimate. The posted messages could be spam. Note that although spam traditionally refers to massive, unsolicited campaigns trying to promote products or services, we do not restrict ourselves to this behavior alone. Rather, we use the term "spam" to refer to unsolicited campaigns that attempt to execute a variety of attacks, including but not restricted to

I) product advertisements
ii) Phishing and
iii) Malware spreading.

In the example of Figure 1(a), user A's account is compromised and sends a spam message to user C, trying to direct user C to some malicious website. Once user C signs in, the spam message will be displayed to him, exposing him to potential threats.

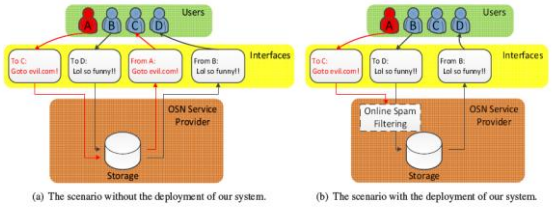

Figure 1: A simplified OSN architecture, only illustrating the components related to the message exchanging functionality. User A represents the account controlled by spammers.

III GOAL

Our goal is to design an online spam filtering system that is deployed at the OSN service provider side, as Figure 1(b) shows. Once deployed, it inspects every message before rendering the message to the intended recipients and makes immediate decision on whether or not the message under inspection should be dropped. In this particular example, the message generated by user A is classified as spam and is thus dropped instantly. The message from user B is legitimate and is stored by the service provider. Later when user C and D sign in to the system, C will not see the dropped Spam message, as spam message will get converted into stars(***).

IV SCOPE OF WORK

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.

REFERENCES

[1] C. Yang, R. Harkreader, and G. Gu, "Die free or live hard? Empirical evaluation and new design for fighting evolving twitter spammers," in *Proc. 14th Int. Symp. Recent Advances in Intrusion Detection (RAID'11)*, Menlo Park, CA, USA, Sep. 2011.

[2] Costolo: Twitter Now Has 190 Million Users Tweeting 65 Million Times A Day, 2010 [Online]. Available: http://techcrunch.com/ 2010/06/08/twitter-190-million-users/

[3] Acai Berry spammers hack Twitter accounts to spread adverts, 2009[Online]. Available: http://nakedsecurity.sophos.com/2010/12/13/ acai-berry-spam-gawker-password-hack-twitter/

[4] New Koobface campaign spreading on Facebook, 2011 [Online].Available: http://forums.cnet.com/7726-6132_102-5064273.html

[5] Twitter-based Botnet Command Channel, 2009 [Online]. Available: http://ddos.arbornetworks.com/2009/08/twitter-based-botnet-command-channel/

[6] Twitter phishing hack hits BBC, Guardian and cabinet minister,2010 [Online]. Available: http://www.guardian.co.uk/technology/ 2010/feb/26/twitter-hack-spreadphishing

[7] A new look at spam by the numbers, 2010 [Online]. Available:http://scitech.blogs.cnn.com/2010/03/26/a-new-look-at-spam-by-thenumbers

[8] A. H. Wang. Don't follow me: Spam detection in twitter. In *Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on*, pages 1–10, July2010.

[9] Twitter On Track For 500 Million Total Users By March. http://www.mediabistro.com/alltwitter/twitter-active-total-users b17655, March 2012