Spammers' Networks within Online Social Networks: A Case-Study on Twitter

Saptarshi Ghosh Gautam Korlam Niloy Ganguly
Department of CSE, Indian Institute of Technology Kharagpur, India
{saptarshi,gautam,niloy}@cse.iitkgp.ernet.in

ABSTRACT

We analyze the strategies employed by contemporary spammers in Online Social Networks (OSNs) by identifying a set of spam-accounts in Twitter and monitoring their link-creation strategies. Our analysis reveals that spammers adopt intelligent 'collaborative' strategies of link-formation to avoid detection and to increase the reach of their generated spam, such as forming 'spam-farms' and creating large number of links with targeted legitimate users. The observations are verified through the analysis of a giant 'spam-farm' embedded within the Twitter OSN.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services; J.4 [**Computer Applications**]: Social and behavioral sciences

General Terms

Experimentation, Security

Keywords

Online Social Networks, Twitter, spam-farms

1. INTRODUCTION

Popular OSNs (e.g. Twitter, Facebook) of today are being increasingly targeted by spammers and other malicious users to promote affiliate websites and disseminate malware. Though many techniques for controlling malicious users have been proposed, such as machine-learning based techniques [4], Sybil-defense schemes [5], and so on, large amounts of spam continue to plague the popular OSNs. This is most probably because the OSN authorities are unwilling to deploy automated methods at large scale to detect suspicious user-accounts and suspend them, fearing that wrong decisions would lead to serious discontentment among users of the OSN. Hence spam-accounts are suspended mostly after a sufficient number of users report them as spam. However, most legitimate users are unwilling to invest time in reporting against spammers, hence spammers are being allowed more time to spread spam in the OSNs.

In our ongoing research, we are investigating methods to overcome this problem by understanding the strategies employed by spammers in the OSNs; knowledge of these strategies can be utilized for designing effective and proactive spam-control techniques. Here we demonstrate that analysis of the *social link-creation patterns of spammers* can provide valuable insights into the spammers' strategies.

A primary objective of spammers in OSNs is to acquire a large number of social links or friends ('followers' in Twitter[1]), since (i) this helps them pose as popular users and thus evade suspicion of spam-detection algorithms, and more importantly, (ii) this enables rapid dissemination of spam to a large audience using the one-to-many communication methods provided in OSNs (e.g. public 'tweets' multicast to all followers in Twitter). In most OSNs, one-to-all-friends modes of communication are reserved for use across existing social links only, hence spammers aim to acquire a large number of social connections (followers in Twitter) and then use these modes of communication.

What strategies do spammers employ in present-day OSNs to attain this objective? Spammers in the *Web* are known to change the link structure of their affiliate web-sites to create 'link-farms' in order to deceive search engines and thus improve the ranking of one or more of these web-pages [1]. Do spammers in *OSNs* adopt a similar technique of linking to one another to create their own 'spam-farms' in order to pose as popular users? We conducted several empirical experiments on the Twitter OSN to answer these questions, as are described below.

2. ANALYZING SPAMMERS' STRATEGIES

We started with 8 spam-accounts in Twitter (obtained heuristically from a large crawl of Twitter users studied by us in [2] and verified manually) and crawled their neighbourhood to detect other suspicious user-accounts. As reported in [4], spammers in Twitter repeatedly post tweets containing URLs of their affiliate web-sites, hence we use the *number of repeated URLs in recent tweets* to identify suspicious users. Note that this feature is *not* a confirmatory test for spammers, since some *promoters* also repeatedly post the same URL to advertise their websites or services [4].

We ran *selective* BFS crawls starting from the 8 spammers in August 2010; among the followers and followings of the user currently being crawled, only the suspected ones were added to the BFS queue to be crawled subsequently. We found 3471 suspicious user-accounts within a distance of two

[1]Twitter is a directed social network where user u 'follows' user v if u intends to receive all 'tweets' (messages) posted by v. In Twitter terminology, u is a 'follower' of v and v is a 'following' of u.

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0637-9/11/03.

'hops' from the set of 8 spammers; out of these, 79 have been suspended by Twitter authorities by November 10, 2010. Since active accounts are suspended by Twitter primarily for spam-activity, we assume that accounts that are suspended were actually spam-accounts. Surprisingly, out of these 79 spam-accounts, many had several thousand followers (and followings) not only from other spammers, but also from thousands of *legitimate* users in the OSN.

How did the spammers acquire so many followers from among legitimate users? We observed that there exist many user-profiles which are followed by (and follow) several of the spammers. To gain a detailed insight into the strategies of spammers, we filtered out the top 50 accounts based on the number of spam-accounts connected with them (out of the 79). Of these 50, 5 were already suspended by Twitter, showing that they were spammers as well. We created a Twitter account and followed the other 45. Within a single day, our Twitter account acquired 33 followers - some were reciprocating follows by the users whom we followed, the others were unsolicited follows by users who were themselves suspicious.

This experiment demonstrates that (i) spammers not only follow other spammers, but also target specific legitimate users who frequently follow back (mostly promoters of some service, who consider it a social etiquette to follow back prospective customers) (ii) spammers monitor the followers of the targeted legitimate users, and begin to follow those who follow the targeted users, hoping to get followed back; possibly this is how spammers discover other potential spammers and collaborate with them (e.g. in growing the set of legitimate users to target).

3. LARGE-SCALE ANALYSIS OF SPAMMERS

To verify our findings, we consider a publicly available snapshot of the Twitter OSN as in July 2009 [3]. We attempted to crawl the profile-pages of a large fraction of the 41.7 million users in the snapshot, and discovered more than 30,000 user-accounts that have been suspended by Twitter since July 2009 [2]. Out of these, we studied the 4491 accounts that had more than 1000 each of followers and followings; this large number of social links created show that these accounts had been active ones, hence they were suspended most probably due to spam-activity. Analyzing the social links created by these spam-accounts, we observe the following:

(i) The 4491 spam-accounts had more than 730,000 directed links among them, confirming the presence of a giant spam-farm having a density of 0.036, several orders of magnitude higher than the density of 8.46×10^{-7} for the entire Twitter snapshot. In fact, it is intriguing that spam-accounts are able to find (and link to) other spam-accounts so frequently, within a social network of the size of Twitter.

(ii) On an average, 4.74% of the follow-links created by these spammers lead to other spammers (within the 4491), and this fraction is as high as 12% for some of the accounts

(iii) Fig. 1 plots the fraction of *reciprocated* follow-edges of these spammers against the number of follow-edges created. Compared to the reported 22.1% reciprocation for the entire Twitter snapshot [3], almost all spammers have much higher reciprocation of the follow-edges they create. This

[2]attempts to crawl the profile-page of these lead to a Twitter web-page showing that the account has been suspended

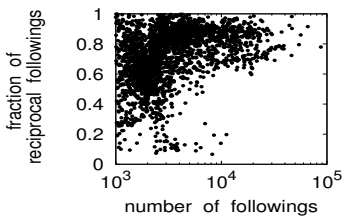

Figure 1: Variation of fraction of reciprocated follow-links with number of follow-links created by spammers

confirms that spammers selectively follow users who follow-back. Moreover, spammers having higher number of followings have higher reciprocation on the average as well, showing that as spammers spend more time in the network and create more links, they are able to filter out more users who follow-back and thus expand the reach of the spam.

(iv) Large overlaps exist among the neighbours of the spammers (e.g. 275 accounts are followed by more than 1000 of the 4491, and 345 accounts follow more than 1000 of the 4491 spammers). implying large-scale collaboration among spammers in identifying potential users to follow.

4. DISCUSSION AND CONCLUSION

We uncover several tactics employed by spammers in OSNs: not only do spammers collaborate among themselves by forming giant 'spam-farms', they also target many specific legitimate users who unwittingly create links with spammers. As a result, though the spammers form dense communities among themselves, such communities get deeply embedded into the social network and become extremely difficult to identify. This can lead to large-scale Sybil-attacks in distributed OSNs, and the Sybil-defense schemes of today - which assume that although a malicious Sybil-attacker can create an arbitrary number of sybil-nodes in the social network, such sybil-nodes can only form a limited number of social links to legitimate (non-Sybil) nodes [5] - are likely to face new challenges in dealing with large spam-farms having millions of links with legitimate users.

Thus we have identified the footprints left by large spam-farms within OSNs, and provided several insights on the link-creation strategies of spammers, that need to be considered while developing anti-spam strategies. For instance, proactive strategies can be tried, such as automatic monitoring of suspicious users in the neighbourhood of reported spammers, and warning legitimate users against reciprocating to 'friend requests' from the monitored suspicious users.

5. REFERENCES

[1] L. Becchetti et al. Link analysis for web spam detection. *ACM Transactions on the Web*, 2:1–42, March 2008.

[2] S. Ghosh et al. The effects of restrictions on number of connections in OSNs: A case-study on Twitter. In *WOSN*, 2010.

[3] H. Kwak et al. What is Twitter, a social network or a news media? In *WWW*, 2010.

[4] K. Lee et al. Uncovering social spammers: social honeypots + machine learning. In *SIGIR*, 2010.

[5] B. Viswanath et al. An Analysis of Social Network-based Sybil Defenses. In *SIGCOMM*, 2010.