Building Artificial Identities in Social Network Using Semantic Information

Kai Chen, Yi Zhou, Li Song, Xiaokang Yang
Department of Electronic Engineering
Institute of Image communication and Information Processing
Shanghai Jiaotong University, Shanghai, China
E-mail: {kchen, zy_21th, song_li, xkyang}@sjtu.edu.cn

Abstract— **as the popularity of social networking sites increase, so does their attractiveness for criminals. In this work, we show how an adversary can build artificial identities using semantic information in social network. Our method make the identities look more like real people, therefore can be used to support many kinds of attacks, such as ASE [1], profile cloning [2]. A prototype of this method is implemented, includes following stages: Firstly, categories of virtual identity are predefined, and each category has multiple properties, such as geographical region, hobby, education, age, interested topic/keywords, etc. Secondly, based on category information, each identity will foster its own "life" semantically, such as edit profile and update status, find hot related news/topic from Google then post to wall, find related groups/networks then request to add in, and find/like/create/comment pages/posts, etc. Thirdly, artificial identity will evolve to multiple stages according to its status (for example, number of friends of real people), single identity with different evolutionary stages is linked together to a group that will help to ensure the number of attack edges [3].**

Keywords- Social Network; Artifical Identity; Semantic Bot

I. INTRODUCTION

Social networking sites have been increasingly gaining popularity. Millions of people daily use social networking sites such as facebook.com, LinkedIn.com. More and more data has been published, at the same time, preserving privacy in social network data becomes an important concern.

Various researchers have shown that social networks can pose a significant threat to privacy of these data. Research work of [4] presented social phishing experiments. They have crawled a number of social networks for Phishing attack. In [1], authors presented a novel automated social engineering cycle which makes traditional social engineering a cheap and attractive attack in social networks. In [2], authors introduced a cross-site profile cloning attack. A new attack that instruments human conversations for social engineering is described in [5]. Research work of [6] introduced an idea of "semantic bot". The European Network and Information Security Agency (ENISA) published a position paper [7] on the Security Issues and Recommendations. SybilLimit presented in [3] is good solution to defense Sybil attacks [8].

In this paper, we propose an identity-build method which can be potentially used as semantic bot for social network. Furthermore, single identity can be organized as group that will help to ensure the number of attack edges (described in [3]) and then make Sybil attack more effective.

II. OUR METHOD

A. Flow of Method

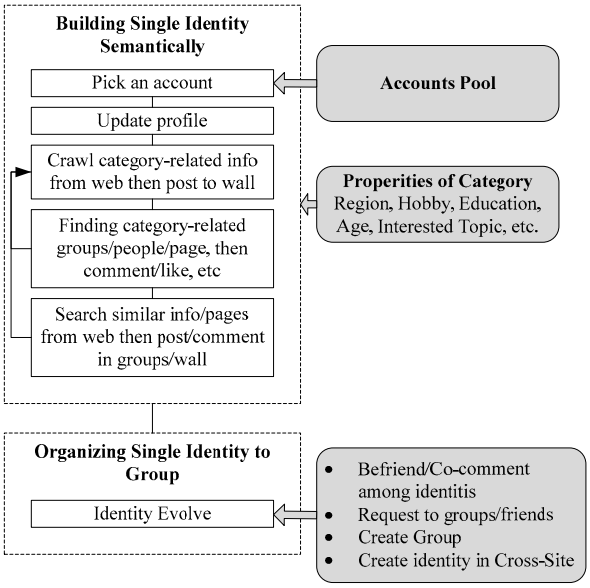

Figure 1, Flow of Method

The flow of our method is shown in Fig. 1.

Categories and their properties are predefined, and part of properties can be fetched automatically, for example, keywords of topic can be analyzed by topic clustering algorithm. Accounts Pool is generated manually, and could be built automatically by attacking CAPTCHA.

Crawler will search to find related info/pages using search engine, such as Google, from web based on topic keywords, and may get more results using recommended search keywords provided by search engine. Furthermore, crawler will search related peoples/groups/pages using topic keywords, and will get more new keywords by clustering such results. Then similar pages from search engine will be posted/commented to pages/groups/peoples as well.

B. Evolution and Group

Single artificial identity can be evolved in different stages according to what we called-"credit points". Currently, the credit points are determined by the number of real-friends,

978-0-7695-4375-8/11 $26.00 © 2011 IEEE
DOI 10.1109/ASONAM.2011.33

the number of post/comments, and the duration of account. There are 4 predefined stages in our preliminary design: (1) Stage-0, a new created identity; (2) Stage-1, identity with 100 posts/comments/like pages; (3) Stage-2, identity with 5 real friends; (4) Stage-3, identity with 30 real friends. Above parameters are set according to human experiences. Current attacks, such as [1, 2, 4, 5], can be used during evolution between stages. In fact, according to [9], even without help of those attack methods, still some personal profile can be read.

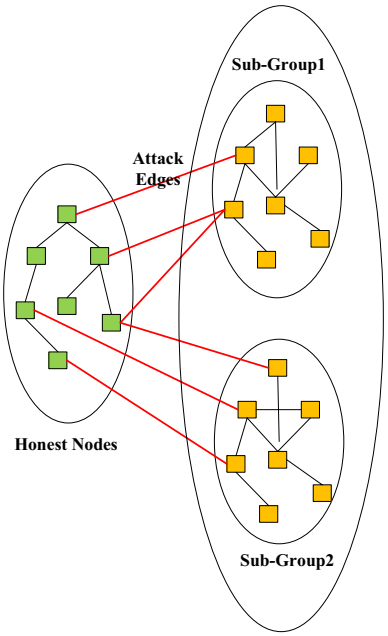

Figure 2, Attack Edges

All the identities with same or similar category can be organized into a sub-group and finally into a group, identities with different stages have their different roles in group, for example, only stage-3 identity can initiate real attacks. As Fig 2 shown, this mechanism can help to ensure the number of attack edges (described in [3]) and then make Sybil attack more effective.

III. EXPERIMENT

We used Linode VPS[1] as hosts. Each Linode VPS is a Linux virtual machine with IP address in different regions, such as London in UK, Newark, Dallas, Atlanta and Fremont in US. In current experiment, each VPS hoses 20 Facebook identities. In the future, due to the excellent extensibility of VPS, more artificial identities can be hosted easily while difficult to be detected by Facebook.

To simulate a web browser, we tried to use chickenfoot[2] as automated Facebook bot in a first attempt. chickenfoot is a kind of web browser extension, therefore, the performance of is limited and not suit for potential large scale bots. Then we implement our bot based on HtmlUnit[3], which models HTML documents and provides an API that allows programmer to invoke pages, fill out forms, click links, and other tasks which people do in "normal" browser. Facebook open APIs[4] are also used to do general crawling jobs, such as search groups/peoples/pages.

In our preliminary experiment, we built total 100 Facebook identities in 5 VPS, posted 2000 messages and comments in 3 days without any block from Facebook.

IV. CONCLUSION AND FUTURE WORKS

In this paper, we propose an identity-build method which can be potentially used as semantic bot for social network. Furthermore, single identity can be organized as group that can make Sybil attack more effective.

Till now, our work is preliminary, there are lot of works are needed toward a mature work. Large scale identities should be built and more experiments should be evaluated by comparisons with start-of-art works, and defending mechanisms will be analyzed as well.

ACKNOWLEDGMENT

The work is partially supported by the National Grand Fundamental Research 973 Program of China with grant No.2007CB316506, No.2010CB731406, 2010CB731401.

REFERENCES

[1] M. Huber, S. Kowalski, M. Nohlberg, and S. Tjoa. Towards automating social engineering using social networking sites. In CSE (3), p. 117–124. IEEE Comp. Soc., 2009.

[2] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All your contacts are belong to us: Automated identity theft attacks on social networks. In WWW '09, p. 551–560, New York, NY, USA, 2009. ACM.

[3] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. SybilLimit: A Near-Optimal Social Network Defense against Sybil Attacks. In IEEE Symposium on Security and Privacy, 2008.

[4] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer, Social phishing, Communications of the ACM, vol. 50, no. 10, pp. 94–100, 2007.

[5] L. Tobias, P. Veikko, B. Davide and K.Engin, Honeybot, your man in the middle for automated social engineering, LEET'10, 3rd USENIX Workshop on Large-Scale Exploits and Emergent Threats, San Jose, 27 April 2010.

[6] R.Temmingh and K. Geers, Virtual Plots, Real Revolution, the Conference on Cyber Warfare 2009.

[7] G. Hogben, Security Issues and Recommendations for Online Social Networks, Position Paper. ENISA, European Network and Information Security Agency, 2007.

[8] J. R. Douceur, The Sybil Attack. In Electronic Proceedings for the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02), March 2002.

[9] D. Rosenblum, What Anyone Can Know: The Privacy Risks of Social Networking Sites, Security & Privacy, IEEE, vol. 5, no. 3, pp. 40–49, May-June 2007.

[1] http://www.linode.com/

[2] http://groups.csail.mit.edu/uid/chickenfoot/

[3] http://htmlunit.sourceforge.net/

[4] http://developers.facebook.com/