2012 ASE International Conference on Social Informatics (SocialInformatics 2012) / 2012 ASE International Conference on Cyber Security (CyberSecurity 2012) / 2012 ASE International Conference on BioMedical Computing

Organizational Intrusion:

Organization Mining using Socialbots

Aviad Elishar, Michael Fire, Dima Kagan, and Yuval Elovici
Telekom Innovation Laboratories at Ben-Gurion University of the Negev
Information Systems Engineering Department, Ben-Gurion University of the Negev, Beer-Sheva, Israel
{aviade, mickyfi,kagandi,elovici} @bgu.ac.il

***Abstract*— In the recent years we have seen a significant growth in the usage of online social networks. Common networks like Facebook, Twitter, Pinterest, and LinkedIn have become popular all over the world. In these networks users write, share, and publish personal information about themselves, their friends, and their workplace. In this study we present a method for the mining of information of an organization through the use of social networks and socialbots. Our socialbots sent friend requests to Facebook users who work in a targeted organization. Upon accepting a socialbot's friend request, users unknowingly expose information about themselves and about their workplace. We tested the proposed method on two real organizations and successfully infiltrated both. Compared to our previous study, our method was able to discover up to 13.55% more employees and up to 18.29% more informal organizational links. Our results demonstrate once again that organizations which are interested in protecting themselves should instruct their employees not to disclose information in social networks and to be cautious of accepting friendship requests from unknown persons.**

I. INTRODUCTION

The last ten years has resulted in a significant growth in Internet usage by users all over the world. This is particularly true in social networks [1]. In recent years, social networks have become an integral part of our daily lives. In most cases, users in social networks must create a profile, and then they are open to creating connections with existing friends as well as connecting with strangers. A user's profile can include their real name or a pseudonym. In many cases, a user's profile also can include photographs, birthday, hometown, religion, ethnicity, and personal interests. In undirected social networks, like Facebook[1] and LinkedIn[2], members connect to others by initiating a friend request that must be accepted by the other party in order to establish a friendship link between the users. In many cases, as a result of the privacy settings of a social network, when a user A accepts a friend request sent by user B, then user B receives the privilege to access user A's profile details and vice versa. Online social network users use the network for a number of purposes. The main motivation for social networking online is to communicate with others and to maintain current relationships. Other popular activities include updating others on their activities and whereabouts, sharing photos and archiving events, getting updates on activities by friends, sending messages privately, displaying curriculum vitae and workplaces, etc. [2].

Today there is a wide range of different social networks available to the public ranging from social networks for academics, such as Academia.edu[3] to Dogste[4], a social network for dog owners. As of October 2012, Facebook is the biggest social network in the world [3] with more than a billion monthly active users. Moreover, 57.8% of Facebook users log onto the site on a daily basis, with an average of more than 7 hours per month spent online and more than 30 billion pieces of content shared each month. Facebook has become so popular that it is a well-known phenomenon for users to share their information with others such as photos, posts, statuses, locations, work places, and many more other personal details [4]. This network phenomenon is not without negative effects which may occur when users expose private and sensitive information about themselves, their friends, and their workplace [4]. The problem was highlighted in 2011 by Boshmaf et al. [5] who evaluated how vulnerable online social networks are to a large-scale infiltration by socialbots. These socialbots were designed to be stealthy, i.e., able to pass themself off as human beings. The goal of these socialbots was to infiltrate users so as to reach an influential position. This position can be then exploited to spread misinformation and propaganda in order to bias the public opinion. Boshmaf et al. [5] demonstrated that when socialbots infiltrate a targeted OSN, they can further harvest private users' data, such as email addresses, phone numbers, and other personal data associated with monetary value. To an adversary, such data are valuable and can be used for online profiling and large-scale email spam and phishing campaigns. Moreover, in 2010, Kwak et al. [6] recounted how they successfully crawled the entire Twitter site and obtained thousands of trending topics, millions of user profiles and tweets, and billions of social relations. In 2012, Fire et al. [7] presented algorithms for constructing organizational crawlers which collected data from the Facebook social network in order to gather public information on users who worked in a specific organization. By using publicly available data only, Fire et al. [7] restructured parts of the targeted organization and discovered hidden departments and leadership roles, among the many discoveries, in each organization. This paper is highly influenced both by the Boshmaf et al. [5] study on infiltration of online social

[1] http://www.facebook.com/
[2] http://www.linkedin.com/
[3] http://academia.edu/
[4] http://www.dogster.com/

978-0-7695-4938-5/12 $26.00 © 2012 IEEE
DOI 10.1109/SocialInformatics.2012.39

networks by socialbots and by the study of Fire et al. [7] where publicly valuable information about an organization's structure was discovered. For our paper, we combined the two approaches and used public Facebook profiles in order to study targeted organizations in different aspects.

II. Related Work

In the last few years, many studies have dealt with several of the main topics of our study. Some of them dealt the privacy problem in social networks and its dangers. Other studies discussed exposed data in social networks which can be gathered, and many studies discussed crawling methods, infiltration to targeted organizations, Facebook defenses, etc.

A. The Privacy Problem in Social Networks

In 2006, Barnes [4], concerned about the potential misuse of personal information in social networks, illustrated how easily young adults give up personal information in order to join social networks on the Internet. Barnes emphasized that people in general - are not aware of the potential dangers which can result in revealing personal information online. Such information can come in the form of home address, phone numbers, pictures, etc. The solution Barnes arrived at is not a simple one; in order to tackle the issues which can result from teens and their loss of Internet privacy, a keen awareness and effort by all levels of the society must be brought about and executed.

In 2007, Dwyer et al. [2] investigated trust and privacy aspects on social networks sites. They compared two social networks, Facebook and MySpace, and their results suggested that in regards to online interactions, trust is not as necessary in the building of new relationships as it is for face to face encounters. The authors illustrated that on online sites, the existence of trust and a willingness to share information do not automatically translate into new social interactions. They further indicate that online relationships can develop on sites where the perceived trust and privacy safeguards are in fact weak.

In 2009, Lindamood et al. [8] argued that some of the information revealed inside social networks is private and it is possible that corporations could use learning algorithms on the released data in order to predict undisclosed private information. They found that removing trait details and friendship links is the best way to reduce classifier accuracy, however this method is not considered feasible when maintaining the use of social networks.

Facebook acknowledges the acute problem of information exposure and has therefore taken steps to protect users from malicious attacks and information gains by activating the Facebook Immune System (FIS). In 2011, Stein et al. [9] described FIS as an adversarial learning system that performs real-time checks and classification on every read and write action on Facebook's database, all for the purpose of protecting its users and the social graph from malicious activities. They further elaborated on the design of the FIS, the challenges FIS faced, etc.

B. Exposed Data in Social Networks and Crawling Methods

In 2007, Chau et al. [10] discussed retrieving information from social networks and their implementation of crawlers for online social networks. Using crawlers, they visited a total of approximately 11 million auction users, around 66,000 of which were completely crawled. In 2007, in an attempt to study the characteristics of online social network graphs in large scales, Mislove et al. [11] examined data gathered from four popular online social network sites: Orkut[5], YouTube[6] , Flickr[7], and LiveJournal[8]. They reached a data set containing over 11.3 million users and 328 million links. In 2010, Kwak et al. [6] crawled the entire Twitter[9] site and obtained 1.47 billion social relations, 41.7 million user profiles, 106 million tweets, and 4,262 trending topics. They analyzed the tweets of the top trending topics and reported that they were able to classify the trending topics based on the active period and showed that the majority (over 85%) of topics are headline or persistent news. Moreover, they revealed that any retweeted tweet can reach an average of 1,000 users no matter what the number of followers is of the original tweet. In 2012, Fire et al. [7] analyzed several organizations of different types by data mining. They were able to locate the employees by Facebook, LinkedIn, Google search[10] results, the company's web page, and other publicly available sources. In order to do so, they designed and built a web crawler which extracted a network of the informal social relationships of the employees of a given target organization. The standard crawlers were found to be insufficient for performing data collection because they collected many irrelevant profiles and skipped Facebook users. In contrast to standard crawlers, the designated web crawler optimized data collection from users associated with a specific group or organization. Using the designated crawler, they were able to collect publicly available data of six well-known hi-tech companies on three different scales. Moreover, this group of researchers evaluated centrality measures in order to uncover leadership roles inside the organization and demonstrated that the organizations' users who received relatively high closeness centrality scores were more likely to hold management positions inside the organization.

C. Clustering Methods for Analyzing Organizations

In 1979, Tichy et al. [12] described a technique for analyzing organizations using a network that included several network structure attributes, such as clustering, centrality, and density. Moreover, they used their framework to present an analysis of two organizations with several hundred employees. In 2002, Krebs [13] studied Al-Qaeda's organizational network structure attributes following the September 11th attacks. They successfully discovered the organization's leader by using the degree and closeness structural properties of vertices. In 2007, Mishra et al. [14] introduced a new criterion for clustering ubiquitous social networks and provided an algorithm for discovering clusters. They indicated that their algorithm succeeded in finding good clusters.

[5] http://www.orkut.com/
[6] http://www.youtube.com/
[7] http://www.flickr.com/
[8] http://www.livejournal.com/
[9] https://twitter.com/
[10] http://www.google.com/

D. Collecting Data Using Socialbots

In 2011, Boshmaf et al. [5] described methods for infiltrating a targeted large scale online social network by using and building socialbots. During their study, they operated their socialbot on Facebook for about eight weeks. They used images of attractive women as their fake profile picture [5] because they claimed that an adversary usually uses publicly available personal pictures with corresponding gender and age-ranges. They collected data related to users' behavior and their results reached three conclusions. First, online social networks, like Facebook, can be infiltrated with a success rate of up to 80%. Second, depending on a user's privacy settings, a successful infiltration can result in privacy breaches where even more users' data are exposed than would have been in a purely public profile. Lastly, in practice, online social network security defenses are not effective enough to detect or stop a large-scale infiltration as it occurs.

In our study, we combine the methods introduced by Fire et al. [7], and Boshmaf et al. [5] by using fake user accounts in order to study more about the hidden organizations of organizations.

III. Methods and Experiments

There are two main goals in this study. First, though social networks are one of the greatest assets to today's organizations, they also become a non-negligible threat to an organization's confidentiality. As important information and personal information has been accidently exposed in the past by employees, organizations should be aware of these threats in order to take preventive measures. Second, an important goal for our research team was to infiltrate organizations through the Facebook social network. Through these infiltrations we are studying the targeted organizations and their employees. In order to infiltrate a targeted organization, several actions, as depicted in Algorithm 1, had to take place. First, for the infiltration process, we had to crawl on the targeted organization's website and gather public information about its employees who have a Facebook user account. For the crawling process, we created a public user account (referred to as P) in Facebook without any friends or friend requests sent by them. We used the Organization Social Network Crawler, introduced by Fire et al. [7] in 2011. This preliminary process is very important in order to identify the targeted users we should send friend requests to. After we finished crawling and gathering intelligence on the targeted organization's employees, we created a Facebook socialbot account for every organization we wanted to infiltrate (referred as I). Before infiltration to a targeted organization, it is essential that I user profile look like a reliable profile of real person. For this reason we added personal properties of real profiles such as adding posts, choosing images, choosing interests, etc. Moreover, in 2011, Boshmaf et al. [5] demonstrated that the more friends a user has, the more likely the user is to accept new friendships. Therefore, in the first stage we suggested friend requests to random profiles who had more than 1,000 friends. After we succeed in increasing the level of authenticity of our socialbot, we were able to move to the next step of the intrusion process. After fifty random users accepted our socialbot friend requests, we started to automatically send friend requests to employees working in the targeted organization. In order to choose the right employees to send friend requests to, we sent friend requests to the employees of the target organization who had the highest number friends. It is important to mention that we conducted preliminarily intelligence which helped us reconstruct parts of the organization's social network and revealed the users with the highest number of Facebook friends. After the socialbot was friends with at least ten of organization's employees, we began to send friend requests to the employees with the highest number of mutual friends with our socialbot. During this intrusion process, we limited our friend requests to no more than 20 requests per day. Moreover, we stopped the intrusion process after our socialbot was blocked by the FIS due to a low friend request acceptance rate. We then collected enough information from the targeted organization.

In order to gain valuable and non trivial insights onto the organizational structure of targeted organization, we decided to create clusters for the public network. We wanted to verify whether we would be able to find more clusters after the intrusion compared to the stage before the intrusion. As a clustering method we used the Markov Clustering Algorithm, (MCL) [14] differing from Fire et al. [7], who used the Girvan-Newman Clustering Algorithm. MCL is a way to cluster a graph by flowing through a network. This algorithm has been widely used for clustering in biological networks. However, it demands that the graph be sparse and therefore we found it problematic for our proposes. Additionally, we also used the closeness centrality measure in order to uncover leadership roles inside the crawled and infiltrated organizations and evaluated the percentage of leadership roles of the top 20 users (precision@20) who received the highest closeness centrality score before and after the infiltration. In their study, Fire et al. [7] evaluated different types of centrality measures in order to uncover leadership roles inside an organization's social network. They demonstrated that an organization's employees with high closeness centrality scores were more likely to hold management positions inside the organization. We maintained their approach by choosing the top 20 employees who received the relatively highest values in the closeness centrality algorithm after removing our socialbot user from the organization's network. We took the 20 users and evaluated how many of them held management positions.

ALGORITHM 1: Socialbot Organizational Intrusion

Input: Input: A set of seed URLs (S) to Facebook profile pages of organizations employees.
Output: A set of Facebook profiles and their connections.
Create P1 in Facebook.
CollectedPublicConnectionGraph ← Organization Crawler(S,P1)
Create I1 in Facebook
while $NumOfFriends \leq 50$ **do**
    SendFriendRequest(RandomUser)
end
while $NumOfOrgFriends \leq 10$ **do**
    SendFriendRequest(OrgEmpMaxNumOfFriends)
end
while $NumOfOrgFriends \leq MaxNumberOfFriends$ **do**
    SendFriendRequest(OrgEmpMaxMutualFriends)
end
return Collected pages

A. Limitations of Process

There are many obstacles that can disturb the process described above. The main obstacle we took into account when

trying to infiltrate the group of employees in an organization was the Facebook Immune System (FIS). FIS was investigated by Stein et al. [9] and is an adversarial learning system that performs real-time checks and classifications on every read and write action on Facebook's database. All these actions occur in order to protect Facebook users and the social graph from malicious activities. For this reason, many elements in the crawling process and in the infiltration process may be disrupted, such as blocking and disabling accounts. Blocking accounts occurs when Facebook recognizes a low percent of accepted friend requests. In response, Facebook may block the account that sends the friend requests or temporarily disable it for a week. Moreover, Facebook checks the IP address of the sender of the friend requests, therefore for our purposes we worked with a different permanent IP for every public user. If Facebook recognizes different IPs for a single user, it may operate a face recognition test, a security measure handled by Facebook. The test includes five questions. Every question includes three photos of one of the public user's friends and six friends' names as options to the answer. If the user fails two times in a row, then the user is blocked. Another security measure created by Facebook is the authentication by SMS (Short Message Service). If a user sends too many friend requests, and most of them have not been accepted, then Facebook may ask the user to verify that they are a real person by providing a real phone number.

B. *Creating Intrusion Profile*

The creation of intrusion socialbot accounts (I1 and I2) was a manual procedure. We chose a common name for each account in an attempt to look familiar to other users. We additionally chose different profile pictures. In contrast to Boshmaf et al., for female users, we chose photos of the backside of the targeted woman in order to make recognition unfeasible. For male users, we chose pictures of cute puppies. In order to create an authentic looking profile, we added interests and other common properties, such as likes to famous singers, beautiful nature cover images, etc. We even added very common posts to the public profile's wall.

IV. RESULTS

We operated three public Facebook user accounts for intrusion: one socialbot intrusion account (I1) for the O1 organization, a second socialbot intrusion account (I2) for the O2 organization, and one more public Facebook user account P who had no friends and only served to provide us with access to other Facebook users' publicly available data. We also created the last profile in order to demonstrate the differences between constructing the targeted organization's social network using publicly available data and constructing the targeted organization's social network using data collected by our socialbots.

A. *Targeted Organizations*

We collected data on the following two organizations:

- O1 organization - an international publicly held company that specializes in software development. According to public sources, the company employs thousands of employees and is located mainly in North America, Europe, and Middle East.
- O2 organization - an international technology company located in the Middle East and Eastern Europe. The company employs thousands of employees all over the world.

B. *Public Facebook Account*

Using our public profile P, we crawled and constructed the two organizational social networks and received the following results:

- O1 organization - we discovered 1,859 informal links of 309 Facebook users who, according to their Facebook profiles, worked in this organization (see Figure 1). By using the MCL clustering algorithm, we uncovered 21 clusters with an average size of 14.476 employees and a maximum size of 191 employees.

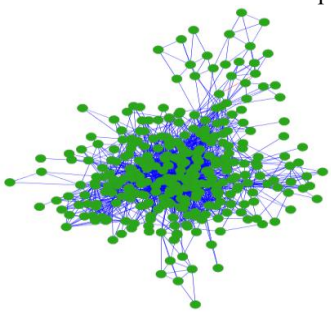

Fig. 1 - O1 social network crawled by P.

- O2 organization - in the crawling process with P on O2, we were able to identify 3,536 informal links of 413 Facebook users who, according to their Facebook profiles, worked in O2 (see Figure 2). By using the MCL clustering algorithm, we uncovered 12 clusters with an average size of 34.167 employees and a maximum size of 319 employees.

C. *Intrusion Profiles (I1, I2)*

Using our two socialbot profiles I1 and I2, we constructed the two organization social network and received the following results:

- O1 organization – I1 socialbot accumulated 57 current O1 employees or past O1 employees from 126 friend requests sent to O1 employees (See Figure 5). The percent of acceptance was 0.452381. By using the crawling process, in I1, we identified 2,199 informal links of 330 Facebook users who, according to their Facebook profiles, worked in this company (See Figure 3). Using the MCL algorithm, we discovered 23 clusters with an average size of 14.217 employees and a maximum size of 254 employees.

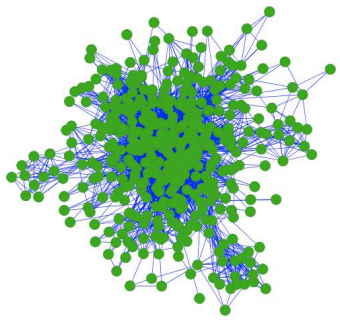

Fig. 2 - O2 social network crawled by P.

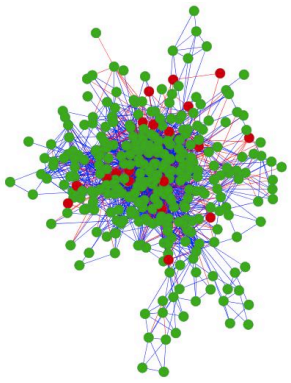

Fig. 3 - O1 social network crawled by I1 (red links represent newly discovered links and red nodes represent newly discovered employees).

- O2 organization – I2 socialbot accumulated 60 current O2 employees or past O2 employees. By using the crawling process, we identified 3,831 informal links of 469 Facebook users who, according to their Facebook profiles, worked in this company (See Figure 4). Using the MCL algorithm, we discovered 25 clusters with an average size of 18.56 employees and a maximum size of 254 employees.

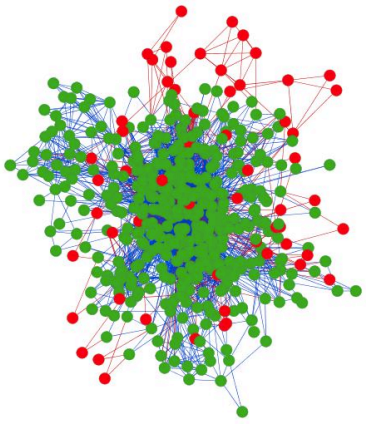

Fig. 4 - O2 social network crawled by I2 (red links represent new discovered links and red nodes represent newly discovered employees).

In the next step we analyzed the network using the closeness centrality analysis (See Table 2). We collected the top 20 employees with the highest closeness centrality scores and evaluated whether these users held leadership positions inside the each organization.

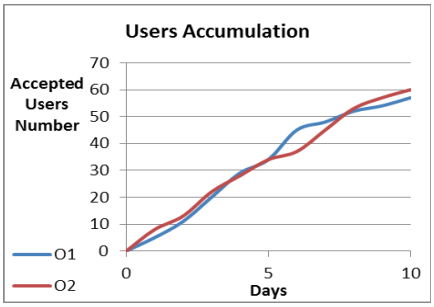

Fig. 5. Users Accumulation

TABLE I. EMPLOYEES AND LINKS BEFORE AND AFTER

Org.	Employees Number / Links Number		
	P	***I1***	***I2***
O1	309 / 1,859	330 / 2,199	-
O2	413 / 3,536	-	469 / 3,831

TABLE II. CLOSENESS CENTRALITY RESULTS

Org.	Employees Number / Links Number		
	P	***I1***	***I2***
O1	0.8333	0.8333	-
O2	0.947	-	0.947

V. CONCLUSIONS

In this study we emphasized an existing privacy problem of organizations in which their employees are members of online social networks. We demonstrated how adversaries who use socialbots can collect employee information in order to reconstruct and better learn the organization's social network. We evaluated our methods by creating two Facebook socialbots which were used to infiltrate groups of employees in two targeted organizations. Our socialbots sent friend requests to Facebook users who work in the organizations we wished to target. By accepting these friend requests, users expose information about themselves and about their workplace. Our method succeeded in discovering up to 13.55% more employees and up to 18.29% more informal organizational (See Table 1) links when compared to the organizations' social network collected by the socialbots. With the MCL clustering algorithm, we were able to uncover more clusters than by using the public profile only. However, by using the closeness centrality measure we received similar results with both the public profile and the socialbots (see Table 2). Our results demonstrate once again that organizations which are interested in protecting themselves should instruct their employees not to

disclose information in social networks and to be careful when accepting friend requests.

We believe this study has several future research directions. A possible direction is to use more sophisticated friend requests algorithms which might receive higher acceptance rates. Another possible direction is to compare different clustering algorithms.

REFERENCES

[1] Ravi Kumar, Jasmine Novak, Andrew Tomkins, "Structure and Evolution of Online Social Networks", 2006.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Catherine Dwyer, Starr Roxanne Hiltz, Katia Passerini, “Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace,” in Americas Conference on Information Systems (AMCIS), 2007.

[3] The New York Times, “Business Day”, July 26-th. 2012.

[4] S. Barnes, “A privacy paradox : Social networking in the United States” Peer-Reviewed Journal On The Internet, 2006.

[5] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, Matei Ripeanu M. Young, "The Socialbot Network:When Bots Socialize for Fame and Money", 2011.

[6] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon, " What is Twitter, a Social Network or a News Media?", 2010.

[7] Michael Fire, Rami Puzis, Yuval Elovici, "Organization Mining Using Online Social Networks", 2012.

[8] Jack Lindamood, RaymondHeatherly, Murat Kantarcioglu, Bhavani Thuraisingham, "Inferring Private Information Using Social Network Data", WWW 2009 MADRID!, 2009.

[9] Tao Stein, Erdong Chen, Karan Mangla, "Facebook Immune System", 2009.

[10] Duen Horng Chau, Shashank Pandit, Samuel Wang, Christos Faloutsos, " Parallel Crawling for Online Social Networks", 2007.

[11] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, Bobby Bhattacharjee, "Measurement and Analysis of Online Social Networks", MCIS(Mediterranean Council for Intelligence Studies) Yearbook, 2007.

[12] Noel M. Tichy, Michael L. Tushman, Charles Fombrun, " Social Network Analysis for Organizations", 1979.

[13] V. Krebs, " Mapping networks of terrorist cells", Connections, Vol. 24, Pages 43-52.

[14] Nina Mishra, Robert Schreiber, Isabelle Stanton, Robert E. Tarjan, " Clustering Social Networks", 2007