International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 18
December 2016

Making Facebook More Secure

P.Srinivas Rao
Associate Professor
Department of CSE,JITS
srithanrao@gmail.com

Dr.Jayadev Gyani
Professor
Department of CSE,JITS
jayadevgyani@gmail.com

Dr.G.Narsimha
Associate Professor
Department of CSE,JNTUHCEJ
narsimha06@gmail.com

Abstract: With 20 million installs a day, third-party apps are a primary motive for the popularity and addictiveness of facebook. Unluckily, hackers have realized the skills of utilizing apps for spreading malware and spam. The crisis is already big, as we discover that at the least 13% of apps in our dataset are malicious. So far, the study group has enthusiastic about detecting malicious posts and campaigns. In this paper, we're going to find that applications are malicious or not. We use expertise collected through observing the posting behavior of common facebook apps which can be going for walks on it. So, first we try to find out the points of malicious apps and different traits of malicious apps which are unsafe to users. In this task, we came up with a framework with which automated detection of false profiles is feasible and is effective. Extra framework makes use of classification methods like support Vector machine, Naive Bayes and selection trees to classify the profiles into fake or actual ones. As, that is an automated detection process, it may be applied without difficulty by means of online social networks which has millions of profiles whose profiles cannot be examined manually.

Key Words: Facebook Apps, Malicious Apps, Profiling Apps, Online Social Networks.

1. INTRODUCTION

In recent days, the social life of everyone has come to be associated with the on-line social networks. These sites have made a radical change in the way in which we pursue our social life. Adding new friends and get in touch with them and their updates have become easier. The online social networks have influence on the science, education, grassroots organizing, employment, trade, and many others. Researchers have been studying these online social networks to see the impact they make on the individuals. Academics can attain the students effortlessly by way of this making a friendly environment for the students to be taught, lecturers now-a-days lecturers are becoming themselves acquainted to these sites bringing online classroom pages, giving homework, making discussions, and so forth. This improves education a lot. The employers can use these social networking sites to hire the individuals who are proficient and in the work, their heritage examine can be carried out quite simply making use of this. Many of the on-line social Networks (OSN) are free however some charge the membership fee and makes use of this for industry purposes and the leisure of them lift money by way of utilizing the promoting. This can be utilized by the government to get the opinions of the public quickly.

The examples of those social networking web sites are sixdegrees.Com, The Sphere, Nexopia which is utilized in Canada, Bebo, Hi5, facebook, MySpace, Twitter, LinkedIn, Google+, Orkut, Tuenti used in Spain, Nasza-

Klasa in Poland, Cyworld traditionally used in Asia, and so on are one of the most widespread social networking websites.

A social network is a social constitution made from a suite of social actors (similar to individuals or corporations) and a collection of the dynamic ties between these actors. The social network perspective presents a suite of methods for analyzing the structure of whole social entities as well as a style of theories explaining the patterns based in these constructions. The study of those structures uses social network evaluation to identify regional and international patterns, locate influential entities, and evaluate network dynamics.

The social network is a theoretical assemble valuable in the social sciences to be trained relationships between members, companies, business, or even complete societies. The term is used to describe a social structure decided by means of such interactions. The ties by way of which any given social unit connects characterize the convergence of the various social contacts of that unit. This theoretical approach is necessarily, relational. An axiom of the social network strategy to working out social interplay is that social phenomena[3] should be certainly conceived and investigated by way of the residences of family members between and within units, rather of the residences of these items themselves. For this reason, one long-established criticism of social network idea is that person company is traditionally unnoticed despite the fact that this is probably not the case in observe. Exactly considering that many types of relations, singular or in mixture type. These network configurations, network analytics are priceless to a broad variety of study businesses.

A number of points of on-line social networks are formed to each and every of the more than300 social networking websites presently in existence. Probably the most characteristic is the capability to create and share a private profile. This profile web page in general entails a photo, some general private information (name, age, sex, and place) and additional space for listing your favorite bands, books, TV shows, movies, hobbies and websites. Most social networks on the internet additionally let you post photos, music, movies and individual blogs to our profile webpage. But the essential purpose of on-line social networks is the capability to seek out and make acquaintance with different website participants. These friends additionally show up as links for our profile page. So visitors can comfortably browse our on-line friend network. Each on-line social network has dissimilar ideas and ways for searching out and contacting potential friends.

Problem statement

MySpace[1] is probably the most open. On MySpace, we're allowed to search for and make contact with persons throughout the complete network, whether they may be far away members of our social network or entire strangers. Nevertheless, we can only achieve access to their full profile information if they conform to become to be our friend and join our network. Facebook, which commenced as a institution social network utility, is way more distinctive and group oriented. On Facebook, which we can simplest seek for people which are in one in every of our established "networks". these networks could comprise the enterprise we're employed for the college we attended,or even our high school. But that we would be able to also become a member of a several of the thousands of similar networks or "groups" which have been created through facebook users,some based on real-life organizations and some that exist simplest in the minds of their founders. Currently, malicious apps[2] often do not include a category, company, or description in their app summary. To detect the malicious facebook applications which may affects to user's private information on his/her profile. As we see user did not get much information about application except name of that application while installing as a result no security available on Facebook.

II.RELATED WORK

1) Detecting and Characterizing Social Spam Campaigns Authors:Hongyu Gao,Jun Hu,Christo Wilson,Zhichun Li,Yan Chen,Ben Y.Zhao.

Description:

In this paper, authors presented an initial study to quantify and characterize spam campaigns launched using accounts on online social networks.

They studied a large anonymized dataset of asynchronous wall messages between Facebook users. We analyze all wall messages received by roughly 3.5 million Facebook users(more than 187 million messages in all),and use a set of automated techniques to detect and characterize coordinated spam campaigns. System detected roughly 200,000 malicious wall posts with embedded URLs, originating from more than 57,000 user accounts. Authors found that more than 70% of all malicious wall posts advertise phishing sites.

They study the characteristics of malicious accounts, and see that more than 97% are compromised accounts, rather than fake accounts created solely for the purpose of spamming. Finally, when adjusted to the local time of the sender, spamming[5] dominates actual wall post activity in the early morning hours, when normal users are asleep.

2) Is the App Safe? A large Sale Study on Application Permissions and Risk Signals Authors:Pern Hui Chia,Yusuke Yamamoto,N.Asokan

Description:

Third-party applications (apps) drive the attractiveness of web and mobile application platforms. Many of these platforms adopt a decentralized control strategy, relying on explicit user consent for granting permissions that the apps request. Users have to rely primarily on community ratings as the signals to identify the potentially harmful and inappropriate apps even though community ratings typically reflect opinions about perceived functionality of performance rather than about risks.With the arrival ofHTML5 web apps,such user-consent permission systems will become more widespread.We study the effectiveness or user-consent permission systems through a large scale data collection of Facebook apps,Chrome extensions and Adroid apps[7].The analysis conforms that the current forms of community ratings used in app markets today are not reliable indicators of priay risks of an app.We find some evidence indicating attempts to mislead or entice users into granting permissions:free applications and applications with mature content request more permissions than is typical:--lookalike applications which have names similar to popular applications also request more permissions than is typical.Authors[4] find that across all three platforms popular applications request more permissions than average.

3) LIBSVM: A Library for Support Vector Machines. Authors: Chih-Chung Chang and Chih-Jen Lin

Description:

LIBSVM is a library for Support Vector Machines (SVMs). Authors have been actively developing this package since the year 2000. The goal is to help users to easily apply SVM to their applications. LIBSVM has gained wide popularity in machine learning and many other areas. In this, authors presented all implementation details of LIBSVM. Issues such as solving SVM optimization problems, theoretical convergence, multi-class classification, probability estimates, and parameter selection are discussed in detail. Support Vector Machines (SVMs) are a popular machine learning method for classification, regression, and other learning tasks. LIBSVM is currently one of the most widely used SVM software.

4) Social Applications: Exploring A More Secure Framework Authors: Andrew Besmer, Heather Richter Lipford, Mohamed Shehab, Gorrell Cheek

Description:

Online social network sites, such as MySpace, Facebook and others have grown rapidly, with hundreds of millions of active users. A new feature on many sites is social applications and services written by third party developers that provide additional functionality linked to a user's profile. However, current application platforms put users at risk by permitting the disclosure of large amounts of personal information to these applications and their developers. This paper formally abstracts and defines the current access control model applied to these applications, and builds on it to create a more secure framework. We do so in the interest of preserving as much of the current architecture as possible, while seeking to provide a practical balance between security and privacy needs of the users, and the needs of the applications to access users' information. We present a user study of our interface design for setting a user-to application policy. The results indicate that the model and interface work for users who are more concerned with their privacy, but we still need to explore alternate means of creating policies for those who are less concerned

BACKGROUND

In this section, we discuss how applications work on Facebook, provide an overview of My Page Keeper[8] (our primary data source), and summary the datasets that we use in this paper

FACEBOOK APPS : Facebook enables third-party developers to offer services to its users by means of Facebook applications. Unlike typical desktop and smart phone applications, connection of a Facebook application by a user does not involve the user downloading and executing an application binary. Instead, when a worker adds a Facebook application to her profile, the user grants the application server: (a) authorization to access a subset of the information listed on the user's Facebook profile (e.g. the user's email address), and (b) permission to perform certain actions on behalf of the user (e.g., the ability to post on the user's wall). Facebook grants these permissions to any application by handing an O Auth 2.0 [4] token to the application server for each user who installs the application. Thereafter, the application can access the data and perform the explicitly-permitted actions on behalf of the user. Fig.1 depicts the steps involved in the installation and process of a Facebook application.

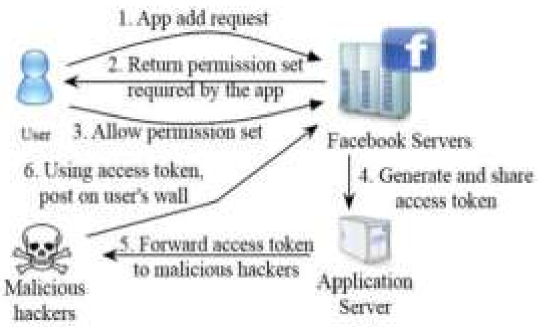

1. App add request
2. Return permission set required by the app
User
3. Allow permission set
Facebook Servers
6. Using access token, post on user's wall
4. Generate and share access token
5. Forward access token to malicious hackers
Malicious hackers
Application Server

Fig 1: Steps involved in hackers using malicious applications to get access tokens to post malicious content on victims' walls.

Operation of malicious applications: Malicious Facebook applications typically operate as follows.

- **Step 1:** Hackers prove users to install the app, usually with some fake promise (e.g., free iPads).

- **Step 2:** Once a user installs the app, it redirects the user to a web page where the user is requested to achieve tasks, such as completing a survey, again with the lure of fake rewards.

- **Step 3:** The app thereafter accesses personal information (e.g., birth date) from the user's profile, which the hackers can hypothetically use to profit.

- **Step 4:** The app makes malicious posts on behalf of the user to lure the user's friends to install the same app .

This way the cycle continues with the app or colluding apps reaching more and more users. Private information or surveys can be "sold" to third parties to eventually profit the hackers.

Mathematical model

Let S is the Whole System Consists:

S = {U, P, Req, A, APP}.

1. U is the set of number of user on the facebook.

U={u1,u2,……….un}.

2. P is the set of number of permission set for user .

P={p1,p2,……………pn}.

3. Req is set of number of add app request from user to server.

Req={a1,a2,……………an}.

4. A is the set of number of set of access tokens of user.

5. APP is the set of number of facebook benign application available on facebooks application server.

APP = {ap1,ap2,………apn}

Step 1: At first user sends request to facebook server for adding an application to his profile like some game app etc.

Step 2: When a request comes to facebook server from client it returns the one set which contains the permissions required to app which he want to install on his profile , permissions like, Application wants to access user information from profile like name, date of birth etc. and this token are send to application server.

Step 3: In this step user allow the access the information from his profile to that particular app, Here user doesn't aware that whether that app is benign or malicious so, here our FRAppE comes in picture. FRAppE checks whether that app is malicious or benign by applying some classifications such as FRAppE Lite and FRAppE.

FRAppE Lite[10]: This is the initial level detection or classifier i.e. FRAppE Lite checks the application ID no, name and location of application and verifies with the available benign application in the application server.

FRAppE: This is actual step of detecting the malicious apps in the facebook. If an application is found malicious then that application will be blocked for all the users so, that in future users don't get request from that application to add.

Step 4: In this step, the FRAppE[9] allows only the benign apps to add on user's wall.

Output: Detecting malicious apps and providing access to only benign apps to user.

III.PROPOSED METHOD

The proposed framework in the fig.2 shows the sequence of actions that need to be followed for continues detection of fake profiles with active leaning from the feedback given by the classification algorithm. This framework can easily be implemented by the social networking sites.

1. The detection process begins with the selection of the profile that needs to be tested.
2. Once the profile is selected,then the required attributes are selected on which the classification algorithm is applied.
3. The extracted attributes are passed to the trained classifier. The classifier gets trained on regular basis as new training data is feed into the classifier.
4. The classifier determines whether the profile is fake or legitimate.
5. The classifier may not be accurate in classifying the profile. So the feedback of the result is given back to the classifier.If the profile is identified as fake, SNS can send notification to the profile to submit identification. If the valid identification is given, feedback is sent to the classifier indicating that the profile was not fake.
6. This process is repeated and as the time proceeds, the no. of training data increases the classifier becomes more and more accurate in predicting the fake profiles.

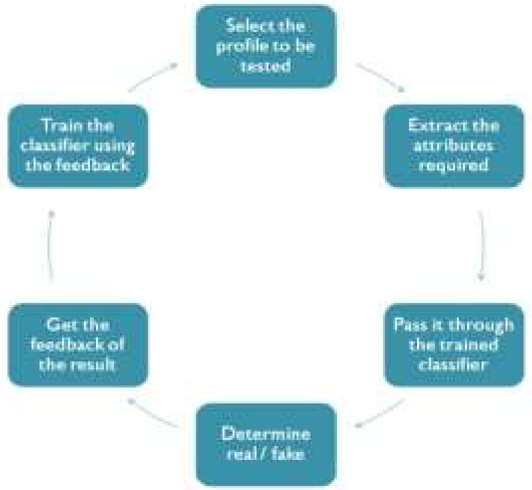
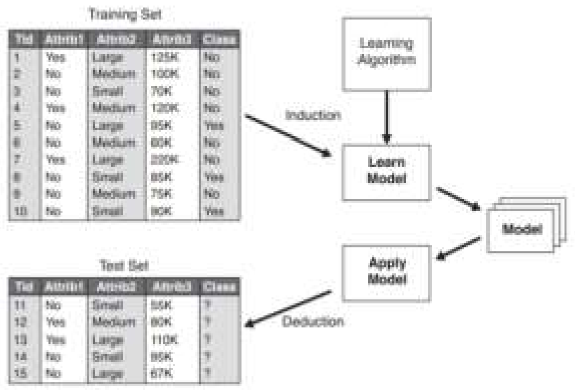

Select the profile to be tested
Extract the attributes required
Pass it through the trained classifier
Determine real / fake
Get the feedback of the result
Train the classifier using the feedback

Fig.2 : Framework for detection of fake profiles and learning

Classification is the process of learning a target function f that maps each records, x consisting of set of attributes to one of the predefined class labels, y. A classification technique is a approach of building classification models from an input data set. This technique uses a learning algorithm to identify a model that best fits the relationship between the attribute set and class label of the training set. The model generated by the learning algorithm should both fit the input data correctly and correctly predict the class labels of the test set with as high accuracy as possible. The key objective of the learning algorithm is to build the model with good generality capability. The fig.3 shows the general approach for building a classification model.



Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Test Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Fig.3: An approach for building a classification model

The classifiers that we have implemented for classifying the profiles are:

- Naive Bayes Classification
- Decision Tree Classification
- Support Vector Machine

IV.RESULTS AND DISCUSSION

Required Datasets:

We need dataset with a mixture of real and fake profiles labeled accordingly. The algorithms need to be trained using the training dataset and should be evaluated using the testing dataset. But there are no such datasets available because of privacy issues.

As there is no standard dataset present, we need to prepare the dataset by scrapping the profiles from facebook. To scrap the data from the profiles, we need to be friends with the profiles which are being scrapped.

A. Data scrapping

Scripts written in python language were used which logs into facebook automatically and scraps required data. Facebook Graph API is also used along with python to extract some required data. Anti scrap detection techniques were implemented to prevent facebook immune system from detecting profiles were scrapped out of which some profiles were hiding data from friends also which were removed from the dataset which left real profiles in the dataset. Barracuda labs is presently working on facebook spam detection making applications for them. They detected and scrapped fake profiles and analyzed the data. We collected the data from them, filtered the profiles in which data is hidden.

B.Attributes Considered

- No. of friends
- Education and work
- Gender
- No. of columns filled in about me
- Relationship status
- No. of photos of the person tagged *
- No. of wall posts posted by the person *
- No. of photos uploaded by the person *

Evaluation parameters

- Efficiency = No. of correct predictions/ Total No. of Predictions
- False Positive rate = No. of real profiles detected as fake/ Total No. of fake profiles to be detected
- False Negative rate = No. of fake profiles detected as real/Total No. of real profiles

From the graph we find that the efficiency of the SVM is highest when the data is well trained and the efficiency of the Naive Bayes is lowest which dont change much when the training dataset increases. As the no. of attributes increases for the training dataset the efficiency of all the algorithms increases. The false positive rate of the SVM is least that means if a profile is detected fake then

the chance of being fake is very high in SVM whereas Naive Bayes shows high false positive rate.

The false negative rate on the other hand is very low for Naive Bayes and the SVM has average false negative rate is the algorithm is well trained.

So, from the results we find that SVM is well suited for classification of the fake profiles in the social networks.

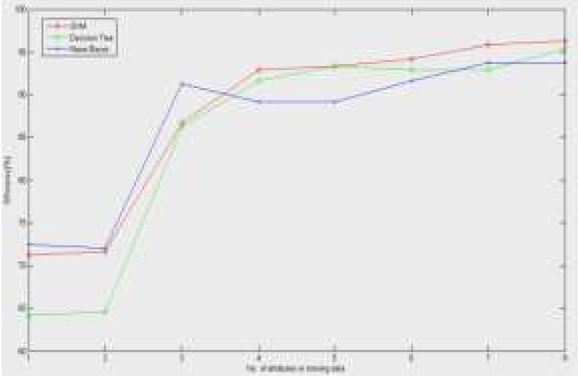


Figure.1: Efficiency vs No. of profiles in training dataset

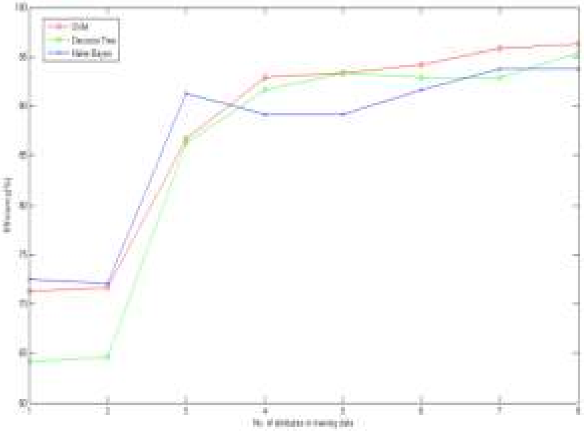


Figure.2: Efficiency vs No. of attributes considered in a profile

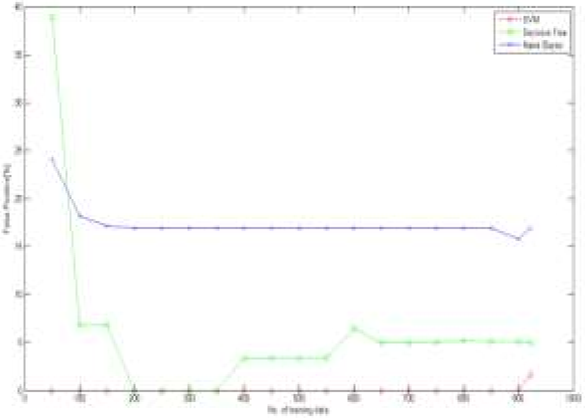


Figure.3: False Positive vs No. of profiles in training dataset

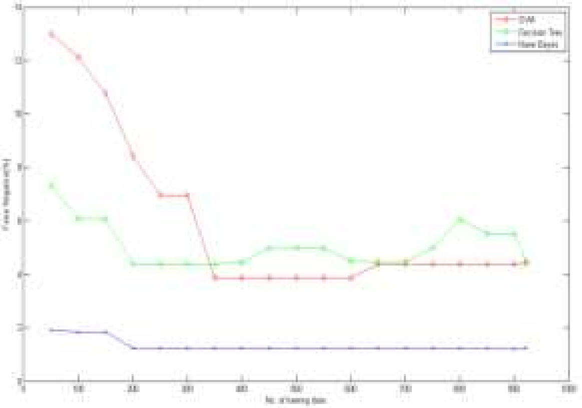


Figure.4: False Negative vs No. of profiles in training dataset

V.CONCLUSION REMARKS

We have proposed a framework using which we can detect fake profiles in any online work networks with a very high efficiency. But our frame work does not address all the characteristics of a profile. Fake profile identification process can be improved by combining Machine learning and NLP techniques .As part of the future work we try to implement this approach.

REFERENCES

[1] T. Stein, E. Chen, and K. Mangla. Facebook immune system. In Proceedings of the 4th Workshop on Social Network Systems, SNS, volume 11, page 8, 2011.

[2] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu. The socialbot network: when bots socialize for fame and money. In Proceedings of the 27th Annual Computer Security Applications Conference, pages 93{102. ACM, 2011.

[3] C. Wagner, S. Mitter, C. K¨orner, and M. Strohmaier. When social bots attack: Modeling susceptibility of users in online social networks. In Proceedings of the WWW, volume 12, 2012.

[4] G. Kontaxis, I. Polakis, S. Ioannidis, and E.P. Markatos. Detecting social network profile cloning. In Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on, pages 295{300. IEEE, 2011.

[5] A. Wang. Detecting spam bots in online social networking sites: a machine learning approach. Data and Applications Security and Privacy XXIV, pages 335{342, 2010.

[6] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B.Y. Zhao. Detecting and characterizing social spam campaigns. In Proceedings of the 10th annual conference on Internet measurement, pages 35{47. ACM, 2010.

[7] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia. Who is tweeting on twitter: human, bot, or cyborg? In Proceedings of the 26th Annual Computer Security Applications Conference, pages 21{30. ACM, 2010.

[8] S. Krasser, Y. Tang, J. Gould, D. Alperovitch, and P. Judge. Identifying image spam based on header and file properties using c4. 5 decision trees and support vector machine learning. In Information Assurance and Security Workshop, 2007. IAW'07. IEEE SMC, pages 255{261. IEEE, 2007.

[9] G.K. Gupta. Introduction to Data Mining with Case Studies. Prentice Hall India, 2008.

[10] Rajan Chattamvelli. Data Mining Methods. Narosa, 2010.09