ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.12

A survey on: Malicious Application and Fake user Detection in Facebook using Data Mining

Akash Kumbhar[1], Meghana Wable[2], Supriya Nigade[3], Komal Darekar[4]
BE Student[1, 2, 3, 4]
Department of Computer Engineering
SEC, SomeshwarNagar, Baramati Pune, Maharashtra, India

Abstract:
Online social networks such as Facebook, Twitter, Hike, Instagram and WhatsApp are popularly used by community peoples for personal and professional use. The main aim of using networking sites are sharing photos, audio and video files and other media also written works, posting a resume for job searching etc. Along these apps Facebook is widely used for communicating past and current friends, for advertising about business, products, textiles, education, etc. Before using these sites, we also focused on personal and professional data security. Hackers well knew the potential of using these apps for spreading malware and spam. In this paper, our contribution aims about detecting unauthorized users and also malicious applications. We, therefore use FRAppE--Facebook Rigorous Application Evaluator--which argued to focus on detecting harmful applications on Facebook. FRAppE also contributes for identifying fake user account to prevent the personal information. Our result indicate that classifying malicious apps and detect fake user requests on Facebook. Our work concludes classifying spam URL's and identifying fake user accounts for security purpose.

Keywords: Online social network, spam, malicious applications, fake user.

I. INTRODUCTION

Online social network is a platform that benefits peoples to create there profiles, finding and making friends. Online social networks are more popular now days. There are some OSN popular networks such as Facebook, Instagram and Twitter. Our key contribution is for Facebook. Now a day's third party apps encourage the enhancement in online social network (OSN). Such enhancements include interesting or entertaining ways of communicating among online friends and diverse activities such as playing games or listening to songs. That is Facebook provides an API [4] that facilitates integration of user experience. 20M apps installed everyday [3] and interesting to know there are 500K apps are available on social networks [5]. Recently, hackers have started taking advantage of the popularity of this third-party apps platform and deploying malicious applications [6]–[8]. Facebook is more popular online social network site among all over the world that causes increase in black market services [23] that encourage growth in fake likes, comments and tags. Fake accounts are categorized into two types called as duplicate accounts and false accounts.

Duplicate Account: A duplicate account refers to an account maintained by a user in addition to his/her principal account.

False Accounts: False accounts are further broken down into two categories user misclassified accounts and undesirable accounts.

- ***User-misclassified accounts:*** It represents the personal profiles created by users for a business, organization, or non-human entity such as a pet (Facebook's terms of service permits such entities as a Page rather than a personal profile).
- ***Undesirable accounts:*** These are the user profiles that are intended to be used for purposes that violate Facebook's terms of service, such as spamming. Fake accounts are mainly used to unfairly increase ones power and influence within a target community [24]. In this work we focused on detecting malicious applications and fake user accounts. Detection of fake user is done on the basis of the user activities and their interaction with other users on Facebook through analysis of user feed data.

There are many ways that hackers can benefit from a malicious app:

1) The app can reach large numbers of users and their friends to spread spam;

2) The app can obtain users' personal information such as e-mail address, hometown and gender; and

3) The app can "reproduce" by making other malicious apps popular. The malicious apps are simplified by using ready-to-use toolkit [9]. Because of this many malicious activities are spent more time on Facebook [10]. Most research on Facebook works in spam and malware [11]-[13]. In this paper, we develop FRAppE, a suite of efficient classification techniques for identifying whether an app is malicious or not and also detection of fake user. For building FRAppE we use My Page Keeper a Facebook's security app [17]. For detection of fake user there are three key contributions of this paper.

- First, it performs data collection on a high restrictive social network, Facebook by capturing all activities (likes, status update, share, etc.) that appear on a user account feed.
- Second, we identify an extensive set of behavioural features which help in identifying the user behaviour.
- Third, number of classifiers are used which uses these features as input to and their performance in regard to fake account detection. The rest of the paper is structured as follows. Next section indicates the related work in brief. In Section II, we describe the literature survey. In section III, we describe the methods and techniques that we use here in this paper. In section IV, we describe the proposed system architecture. In section V, we discuss the conclusion and future

work.

II. LITERATURE SURVEY

1] A technique for computer detection and correction of spelling errors. [18]
Authors: - F. J. Damerau.
Description: A technique for computer detection and spelling errors. This paper describes that which word cannot be match in a dictionary, missing or extra letter or a single transpositions. The unique word which is get entered is compared to the dictionary again, testing each time to see whether the words match- assuming one of these errors occurred. The words which might be wrong or missing are get detected and correct to it.

2] LIBSVM: A library for support vector machine. [19]
Authors: - C.-C. Chang & C.-J. Lin.
Description: The goal of this is to help users to easily apply SUM to their application. LIBSVM (library for support vector machine) has used widely and popularly in machine learning and many other areas. It is use for optimization problem theoretical convergence multicast classification.

3] Beyond Blacklist: learning to detect malicious web sites from suspicious URL's. [20]
Authors: - J. Ma, L. K. Saul, S. Savage, and G. M. Voelker.
Description: In this paper we describe an approach to this problem based on automated URL classification, using statistical methods. The resulting classifier obtain 95-99% accuracy, detecting large number of malicious web sites from their URL's, with only modest false positive.

4] Detecting suspicious URL's in Twitter stream. [22]
Authors: - S. Lee, J. Kim.
Description: Twitter can suffer from malicious tweets containing suspicious URLs for spam, phishing, and malware distribution. Attackers have limited resources and thus have to reuse them; a portion of their redirect chains will be shared. We focus on these shared resources to detect suspicious URLs. We have collected a large number of tweets from the Twitter public timeline and trained a statistical classifier with features derived from correlated URLs and tweet context information. Our classifier has high accuracy and low false-positive and false negative rates.

5] Design and evaluation of real time URL spam filtering service. [21]
Authors: - K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song.
Description: In particular, we find that spam targeting email qualitatively differs in significant ways from spam campaigns targeting Twitter. We explore the distinctions between email and Twitter spam, including the abuse of public web hosting and redirector services. Finally, we demonstrate Monarch's scalability, showing our system could protect a service such as Twitter--which needs to process 15 million URLs/day--for a bit under $800/day.

Problems with existing system:

Hackers realize the potential of spreading malware and spam using malicious apps. Black market services causes growth in fake account. In existing system we can only detect the malicious app. To overcome this problem we developed proposed system.

III. METHODS AND TECHNIQUES

Our Facebook dataset are monitored by MyPageKeeper which is our security application for Facebook. Basically the applications scan your wall and news feed and any time it detects something that looks malicious, it notifies you and inverts you to remove contents. MyPageKeeper scans each URL using machine learning based classifiers that classifies social context associated with URL. MyPageKeeper has false positive rate is 0.005% and false negative rate is 3%. For Implementation and accuracy of My Page Keeper we refer interested readers to [11]. In our work, we need to store data of malicious applications and also need to store data for profiling of users to know which user is fake.

1, Data Collection Methodology

- ***D-Sample Dataset:*** To identify malicious Facebook applications in our dataset, we start with a simple heuristic: If any post made by an application was flagged as malicious by MyPageKeeper, we mark the application as malicious. For every malicious app in the D-sample dataset we consider the time at which we observed the first post made by this app as the time at which the app was launched.
- ***D-Summary Dataset:*** To select an equal number of benign apps from the initial D-Total dataset, we use two criteria:

1) None of their posts were identified as malicious by MyPageKeeper, and
2)They are "vetted" by Social Bakers [17], which monitors the "social marketing success" of apps.
To match the malicious apps from total number of apps i.e. from D-sample dataset it is used.

- ***D-inst Dataset:*** App Permissions: We also want to study the permissions that apps request at the time of installation. For every application App_ID, we crawl https://www. facebook.com/apps/application.php?id= App_ID, which usually redirects to the application's installation URL. Also for in fake user detection we need to collect data related to that account just like facebook real user and fake user.

2. Future Identification:
After collecting all user related data next step is identifying and retrieving set of features from that data attributes. Future Identification retrieves the relationships among all data attributes and used to demonstrate between fake user and real user.

3. Learning Classifiers:
It is a final stage of determination of fake accounts and malicious applications by using supervised machine learning classification algorithm. Supervised learners take datasets as input and construct predictive model. In machine learning we uses algorithms i.e. KNN algorithm and K-MEAN algorithm.

KNN Algorithm:
KNN is K- Nearest Neighbour algorithm. In KNN algorithm it classifies the data as nearest element according to its Euclidian formula.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

The Euclidian distance between two points or tuples, say, x= (x1,x2,…….,xn)
y= (y1,y2,……..yn)

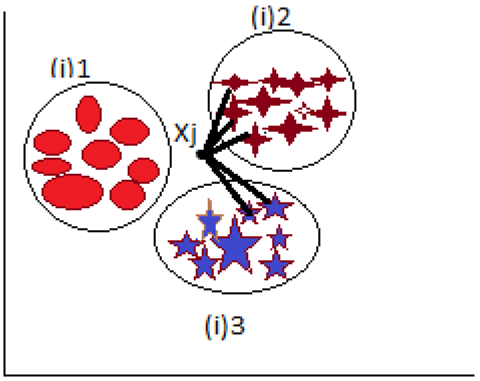

Figure.1. KNN Algorithm K-MEAN Algorithm:

K-MEAN is a clustering algorithm. in k mean algorithm same type of elements are grouped together. In which each cluster is represented as a centre of cluster.

Algorithm:-

1) Calculate mean value of cluster.
2)Assign each item to cluster which has closest mean.
3) Repeat 1 and 2 until we get same mean.
Note:- Initially select 2 random values.

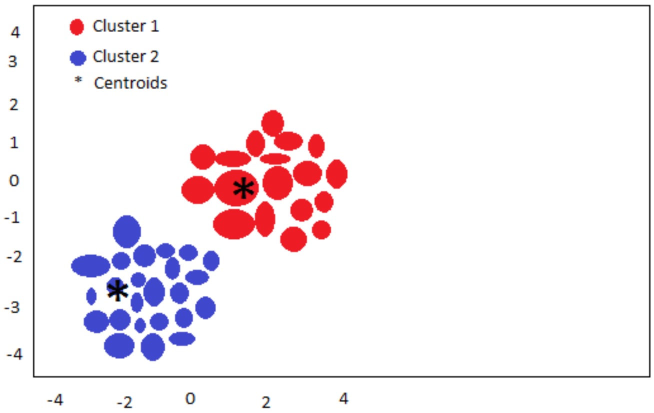

Figure.2. K-Mean Algorithm Input:- K= No. of clusters.
D={t1,t2,t3,………..,tn}
Output:- K= set of clusters.

IV. PROPOSED SYSTEM

Third party apps provide interesting features that means addictiveness in Facebook application. It has several disadvantages like hackers uses its potential for spreading malware and spam. Also due to black market services causes growth in fake user accounts.

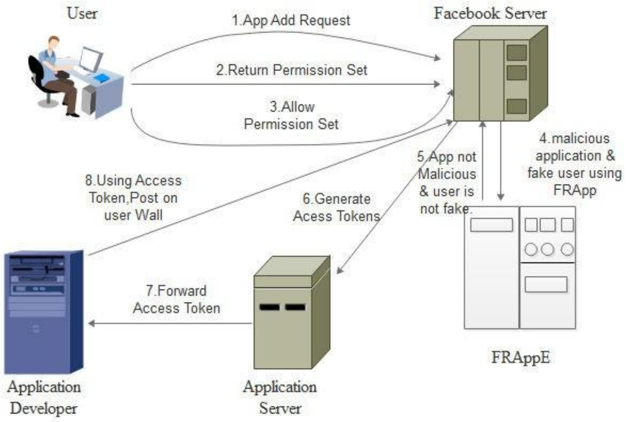

Figure.3. Proposed System Architecture

In proposed system we develop tool which detect the malicious application as well as fake user in facebook. In facebook when user trying to download some application then FRAppE detect that the app is malicious or not and also we detect the fake user present on facebook. First we see how the malicious applications are work: Unlike typical desktop and smart phone applications, installation of a Facebook application by a user does not involve the user downloading and executing an application binary. Instead, when a user adds a Facebook application to her profile, the user grants the application server:
1)Permission to access a subset of the information listed on the user's Facebook profile (e.g.,the user's e-mail address)
2)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 handling an OAuth 2.0[15] token to the application server for each user who installs the application. For detection of fake accounts on a very popular (and difficult for data collection) online social network, Facebook. Key contributions of our work are as follows. The first contribution has been collection of data related to real and fake accounts on Facebook. Due to strict privacy settings and ever evolving API of Facebook with each version adding more restrictions, collecting user accounts data became a major challenge. Our second contribution is the use of user-feed information on Facebook to understand user profile activity and identifying an extensive set of 17 features which play a key role in discriminating fake users on Facebook with real users. Third contribution is the use these features and identifying the key machine learning based classifiers who perform well in detection task out of a total of 12 classifiers employed[25].

Operation of our proposed system:

- **Step 1:** Hackers convince users to install the app, usually with some fake promise (e.g., free iPads).
- **Step2:** Once a user installs the app, it redirects the user to a Web page where the user is requested to perform tasks, such as completing a survey, again with the lure of fake rewards.
- Step3: The app thereafter accesses personal information (e.g. birthdate) from the user's profile, which the hackers can potentially use to profit.
- Step4: The app makes malicious posts on behalf of the user to lure the user's friends to install the same app (or some other malicious app, as we will see later).
- Step5: For fake user it admits the user request into FRAppE. It collects all data according to that account.
- Step6: After that it shows the result for malicious apps and fake user. It prevents the Personal information or surveys to sold it to third parties [16] to eventually profit the hackers.

V. CONCLUSION AND FUTURE WORK

We conclude that applications provide convenient way for spreading malware and spam. Also black market services causes growth in fake accounts. Therefore we contribute our work for facebook.

Future Scope:

(1)Feature Set Improvement:
We plan to refine our existing feature set to incorporate inter-user behavioural pattern with an aim to further improve the accuracy of fake accounts detection.

(2)Improving Detection Accuracies:
We also intend to apply more classification algorithms on larger datasets to further improve our detection accuracies.

(3)Improving Labelling Criteria:
We would like to design and test alternative labelling techniques for the user accounts

VI. REFERENCES

[1]. “Detecting Malicious Facebook Applications” Sazzadur Rahman, Ting-Kai Huang,Harsha V. Madhyastha, and Michalis Faloutsos, http://www.ieee. Org/publications_ standards/ public ations /rights /index.html

[2]. “Towards Detecting Fake User Accounts in Facebook”, Aditi Gupta and Rishabh Kaushal.

[3]. C.Pring,“100socialmediastatisticsfor2012,”2012[Online]. Available:http://thesocialskinny.com/100-social-media-statistics-for-2012/

[4]. Facebook, PaloAlto, CA, USA, Facebook Opengraph API,” [Online]. Available: http://developers. facebook.com/ docs/ reference/api/

[5]. “Wiki: Facebook platform,” 2014 [Online]. Available: http://e n. wikipedia.org/wiki/Facebook_Platform

[6]. “Pr0file stalker: Rogue Facebook application,” 2012 [Online].Available:https://apps.facebook.com/mypagekeeper/?status=scam_report_fb_survey_scam_pr0file_viewer_2012_4_4

[7]. “Whiich cartoon character are you—Facebook survey scam,” 2012 [Online]. Available: https://apps.facebook .com/ mypage keeper/? status=scam_report _fb_survey_s cam_ whiich_ cartoon_character_are_you_2012_0 3_30

[8]. G. Cluley, “The Pink Facebook rogue application and survey scam,” 2012 [Online]. Available: http://nakedsecurity. sophos.com/2012/02/ 27/pink-facebook- survey-scam/

[9].R.Naraine,“Hackersselling$25toolkittocreatemaliciousFacebook apps,” 2011 [Online]. Available: http://zd.net/g28HxI

[10].HackTrix,“StayawayfrommaliciousFacebookapps,”2013[Online]. Available: http://bit.ly/b6gWn5

[11]. M. S. Rahman, T.-K. Huang, H. V. Madhyastha, and M. Faloutsos, “Efficient and scalable socware detection in online social networks,” in Proc. USENIX Security, 2012, p. 32.

[12]. H. Gao et al., “Detecting and characterizing social spam campaigns,” in Proc. IMC, 2010, pp. 35–47.

[13]. H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary, “Towards online spam filtering in social networks,” inProc. NDSS, 2012.

[14].“MyPageKeeper,”[Online].Available:https://www.facebook.com/apps/application.php? Id=167087893342260

[15]. Facebook, Palo Alto, CA, USA, “Application authentication flow using OAuth2.0,”[Online]. Available: http:// d evelopers.facebook.com/docs/authentication/

[16].“11 million bulk email addresses for sale—Sale price $90,” [Online]. Available: http://www. allhome based. com/ Bulk Email Addresses.htm

[17]. Social Bakers, “Social Bakers: The recipe for social marketing success,” [Online]. Available: http://www. social lbakers. Com/

[18] F. J. Damerau, “A technique for computer detection and correction of spelling errors, “Commun.ACM, vol.7, no.3, pp.171–176, Mar.1964.

[19] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines, Trans.Intell.Syst.Technol” vol.2, no.3, 2011, Art.no.27.

[20] J. Ma, L. K. Saul, S. Savage, and G. M. Volker, “Beyond blacklists: Learning to detect malicious Web sites from suspicious URLs,” in Proc. KDD, 2009, pp. 1245–1254.

[21] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, “Design and evaluation of a real-time URL spam filtering service,” in Proc. IEEE Symp. Security Privacy, 2011, pp. 447–462.

[22] S. Lee and J. Kim, “WarningBird: Detecting suspicious URLs in Twitter stream,” in Proc. NDSS, 2012.

[23] B. Viswanath, et al. Towards Detecting Anomalous User Behavior in Online Social Networks. Proceedings of the 23rd USENIX Security Symposium (USENIX Security), August, 2014.

[24] Z. Yang, et al. Uncovering social network sybils in the wild. Transactions on Knowledge Discovery from Data (TKDD), Vol. 8, No. 1, 2014.

[25]” A Survey Paper on FRAPPE – Facebook Rigorous Application Evaluator”. Nitya Sree P1, Sajitha R Swathi2, Vishwanatha3