IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 4, April 2019

Colluding Identity Clone Prediction using Machine Learning

Aashrith B Arun[1], Ajay Kumar D[2], Anilkumara D N[3], Mohammed M Rehan Ahmed[4], Ramesh G[5]

Student, Computer Science and Engineering, The National Institute of Engineering, Mysore, India[1,2,3,4]

Assistant Professor, Computer Science and Engineering, The National Institute of Engineering, Mysore, India[5]

Abstract: As Online Social Networks such as Facebook, Linkedln and Twitter are increasing becoming a part and parcel of one's daily lives, personal information is at stake. Easy access to personal information has made the attackers to steal information from influential users using various forms of attacks. Attackers take advantage of the user's trustworthiness when using Online Social Networks. Hence, there is a need for the third party applications of various Online Social Networks sites to provide defence mechanisms against adversaries. Colluding attack is a way of creating fake profiles of friends of the target in the same OSN or others. Colluders impersonate their victims and send friend requests to the target with an intention to infiltrate their private circle to steal information. These types of attacks are difficult to detect in because multiple malicious users may have a similar purpose to gain information from their targeted user. In this regard, the work intends to overcome this type of attack by addressing the problem of identity clones across multiple Online Social Networks using machine learning.

Keywords: Identity Clone Attack, Machine Learning, Predictive FP growth, Online Social Networks

I. INTRODUCTION

The rapid growth of Web and applications on it has made social networking an important one. Millions of people around the world are connected to each other by Online Social Networks (OSN) which uses web 2.0 technologies. A social networking site is a website where every user has its own profile and has contacts with its friends, family members, employees, share their updates and join new communities and groups which belong to user's interests. Most of the people interact with those persons on OSN, whom they have contact in offline world.

Online Social Networks are a huge store house for massive amount of sensitive and private information of users and their conversations. As the content of information is increasing exponentially on the OSN, protecting the information against hackers, spammers, identity cloning, social bots, phishing become important. But many online users are not aware of privacy schemes and they often reveal a huge amount of personal data on their profiles that can be seen by anyone on the network. So many third party applications of various OSN sites have proposed and employed defence mechanisms against adversaries. Identity Clone attacks (also known as Profile Cloning attacks) is an attack where attacker deceives the user's friend to make healthy relationship with him / her by replicating the user's identity either in the same network or other network.

In this paper, the content based technique is adopted to predict Identity Clone Attack (ICA). The text - based attribute similarity approach is used to find the matching user profiles in a cross site environment. The observed matching user profiles in one OSN is compared with other OSN based on the latest information shared by the user. Users in OSNs are distinguished from one another by the Area Of Interest (AOI) using clustering technique.

II. RELATED WORK

Georges A. Kamhoua et.al [1] proposed to overcome attacks with a combination of both classical and modern threats by addressing the problem of matching user profiles across multiple OSNs by extracting both features and text from a user's profile and building a classifier based on supervised learning techniques. Simulation and experimental results were provided to validate the accuracy of their findings.

Fatemeh Salehi Rizi et.al [2] proposed a new approach for detecting clone identities is proposed by defining profile similarity and strength of relationship measures. According to similar attributes and strength of relationship among users which are computed in detection steps, it will be decided which profile is clone and which one is genuine by a predetermined threshold. Finally, the experimental results are presented to demonstrate the effectiveness of the proposed approach.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 8, Issue 4, April 2019

Michael Fire et.al [3] proposed a thorough review of the different security and privacy risks, which threaten the well-being of OSN users in general, and children in particular. In addition, they present an overview of existing solutions that can provide better protection, security, and privacy for OSN users. They also offer simple-to-implement recommendations for OSN users, which can improve their security and privacy when using these platforms.

Leyla Bilge et.al[4] investigate how easy it would be for a potential attacker to launch automated crawling and identity theft attacks against a number of popular social networking sites in order to gain access to a large volume of personal user information. The first attack they presented is the automated identity theft of existing user profiles and sending of friend requests to the contacts of the cloned victim.

Deepti Dave et.al [5] present a brief knowledge about the attacks and defense mechanisms which are prominent on Online Social networks. They also explains the work which had been performed in the field of detecting clone profiles and cross site clones on OSN. Content related approaches exploits user generated information and is quite simple for detection. Whereas content free method, makes use of information which is stored on OSN server. Hence, it is seen that content free detection approach is more efficient than content based, resulting in less false positive profiles.

Megha Renuka Prasad et.al [6] proposed a mechanism to recognize the fake profiles according to the frequent user activities which exclusively identifies the members' area of interest. To filter the impersonation of member accounts from OSNs, the component accomplishes a classifier to implement adaptable content dependent filtering rules. The adaptability of the component as far as filtering alternatives is upgraded through the expectation of member activities example of a part. In that light, the results are presented from a study regarding the type and amount of information exposed by social network users which allows an invader to clone a profile also supports us in identifying the clone.

III. PROPOSED METHODOLOGY

The proposed system aims to find out if users' friend request can be considered as clones and colluders accounts. Our work takes into account colluding attackers who recreate social circles of the targeted user to further confuse the target and send multiple friend request to the same user. The approach followed is machine learning. The content-database keeps the content associated with area of interests of every member. The postings database tracks all the client activities and the frequencies of activities, consecutively. The member database contains the data of the client. Fig 1 shows the system architecture for the proposed work.

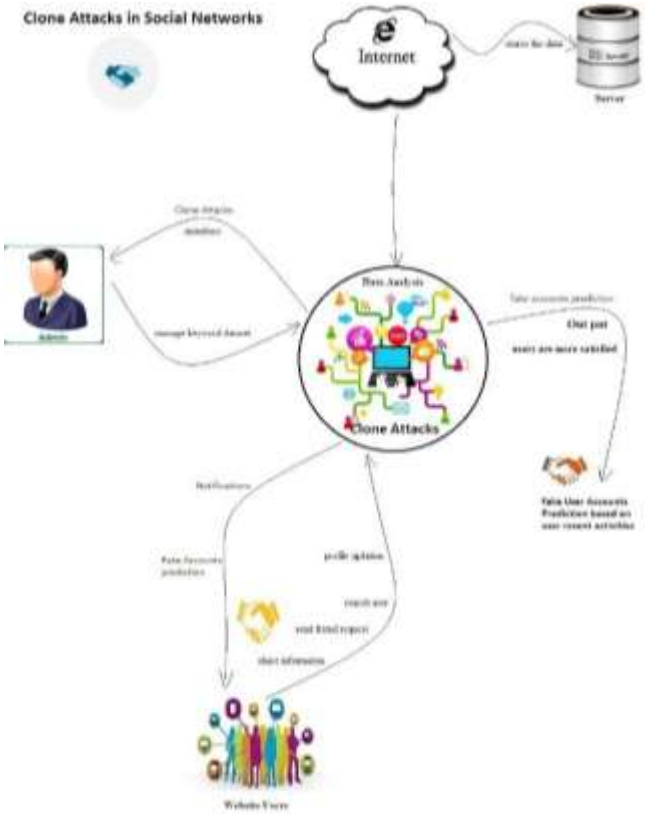

Fig. 1. System architecture

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 8, Issue 4, April 2019

IV. IMPLEMENTATION

First the similarity index is computed to classify the cloned profile and then the member's frequent activities are tracked. Cosine Similarity Method is used to determine how similar the documents are irrespective of their size. The FP-Growth Algorithm is used for mining the complete set of frequent patterns for improving the performance. It uses a divide-and-conquer strategy.

Algorithm:
Input – User Profile and Users Activities
Output – Identification of User's Patterns and Fake profile prediction.

A. User Profile Matching using Fuzzy Sim:

1: Extract the constraints of a user in OSN1 and OSN2.
2: Calculate the similarity index using cosine similarity method

$$Sim(OSN1, OSN2) = \frac{\text{No. of Constraints matching}}{\text{Total no. of constraints}}$$

if (val > Threshold val)
3: Consider those constraints and check if the constraints matching count is greater than minimum support
if (count > support)
Then, Genuine. Else, Fake
4: If Genuine, then track the user activities

B. Tracking the User's Activities: Constrained clustering is a class of semi-supervised learning algorithms. To distinguish between the users in the OSNs we recognize the area of interest of each user by constraint-based clustering technique. We consider three sets: users/members (U), activities (A) and Area of interests (I).
where,
U = {u| users in OSN}={u1,u2...un}
A = {a| activities of users in OSN}={a1,a2...an}
I = {i| subset of activities any user in OSN}

Scan the User (OSN1 and OSN2) in the Database

OSN 1
Step 1: Retrieve shared information from the database (User1 - OSN1)
Step 2: Perform Tokenization, a process of replacing sensitive data with unique identification symbols that retain all the essential information about the data without compromising its security.
Step 3: Clustering the information shared by the users (grouping of similar objects i.e., Area of Interests) by comparing with the predefined dataset (created by the admin).
Step 4: Check if (matching count > minimum_support) [number of contents to compare]
Step 5: Identify the user (OSN1) area of interest (priority wise). (Cluster [AOI] with more number of objects)

OSN 2
Step 1: Retrieve shared information from the other database (User2 - OSN2)
Step 2: Trace user (OSN2) AOI, using the following steps:
Step 3: Tokenization is performed
Step 4: Clustering the messages shared by the users (grouping of similar objects i.e., Area of interests) by comparing with the predefined dataset (created by the admin)
Step 5: Check if (matching count > minimum_support) [number of contents to compare]
Step 6: Compare the present user AOI (OSN1) with the previous user AOI (other OSN2).
If (AOI matches) then
Genuine
Else,
Fake

V. RESULT

For the proposed work, a model dataset having the comparable highlights of association and connections like an ordinary social networking site has been considered. Table 1 shows the member related constraints.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 4, April 2019

Table I : Member related constraints

Sl no.	Constraints	Possibilities
1.	Email ID	Unique ID for a member
2.	First Name, Last Name	Name
3.	Gender	Female, Male
4.	Date of Birth	Day, Month, Year
5.	Education information	School, College, Employment
6.	Location	City, State, Nationality
7.	Language	English, Hindi, Native Language
8.	Contact Number	Phone number
9.	Content type (Area of Interest)	Education, Sports, Art, Music etc.

The results of this paper is depicted in terms of accuracy and graphical representation. The accuracy shows the contrast between the algorithm classification and admin entered classification. The range of the dates should be entered as an input and all the accounts created between the dates is shown with their status. This is shown in Fig 2.

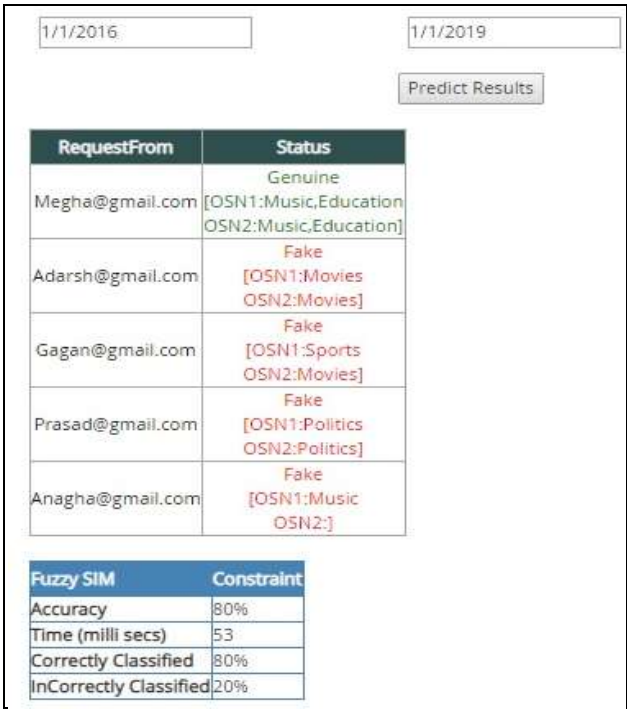

Fig. 2. The output diagram of the accuracy option shown with the table of accounts.

The graph represents the total number of accounts created in the specified range of dates versus the number of fake accounts predicted by the algorithm. This is indicated in Fig 3.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 4, April 2019

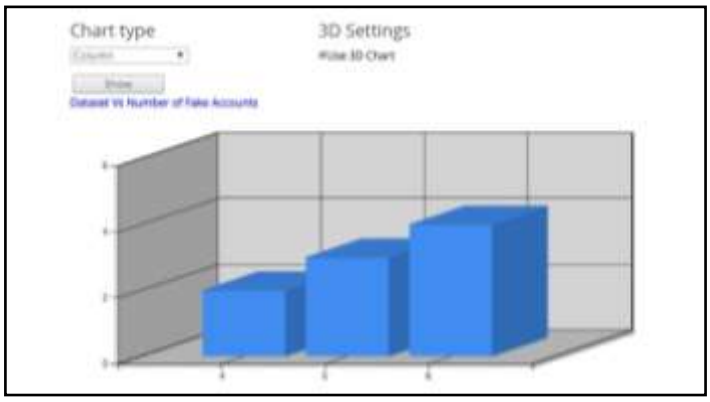

Fig. 3. The output diagram of the graphical option shown with the table of accounts.

VI. CONCLUSION

This paper discusses a predictive mechanism to solve the problem of colluding in Identity Clone Attack in OSNs. We propose a learning model to match the colluding users. The proposed technique has the advantage of being always practical because most of the users have their profiles information public available on OSNs. We propose a predictive colluding identity cloning technique that is effective than the existing ones for the reasons - the usage of Fuzzy Sim and a friend similarity measure to detect colluders in friend requests. Finally, we used precision and recall, to show the quality of the output of our classifier.

VII. FUTURE ENHANCEMENTS

- Adding a feature of SMS verification for the new users of the website during registration.
- Installing an SMS alert module as a future enhancement to the application, where the website user receives an SMS alert for any notifications such as friend request, messages etc.
- Adding a module called Photo Album, where website users can see all photos.
- Introducing the Email Module as a future enhancement to the application where the website user gets an email for any notifications such as friend request, messages etc.

REFERENCES

[1]. Georges A. Kamhoua1 et.al, " Preventing Colluding Identity Clone Attacks in Online Social Networks", 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops, pp 187-192
[2]. Fatemeh Salehi Rizi et.al, "A New Approach for Finding Cloned Profiles in Online Social Networks.", Int. J. of Network Security, Vol. 6, April 2014, ACEEE, pp 25-37.
[3]. Michael Fire et.al, "Online Social Networks Threats and Solutions", IEEE Communications Surveys & Tutorials (Volume: 16 , Issue: 4 , Fourth quarter 2014) , 02 May 2014 ,IEEE, pp 2019 - 2036.
[4]. Leyla Bilge et. al, "All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks.", Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009.
[5]. Deepti Dave et.al , "Detection Techniques of Clone Attack on Online Social Networks: Survey and Analysis" , Elsevier publications, pp 179-186, 2013
[6]. Megha Renuka Prasad et. al, "Advance Identification of Cloning Attacks in Online Social Networks", International Journal of Engineering & Technology, 7 (3.10) (2018) 83-87