IDENTIFICATION OF CLONE ATTACKS IN SOCIAL NETWORKING SITES

Nischitha G[1], Mrs. Sheela N[2]

[1]M.tech Student Dept CS &E, JSS S &T U Mysuru.
[2]Assistant Professor Dept CS & E, JSS S&T U Mysuru.

Abstract- *Social Networking Sites(SNS) is one of the most commonly used online platform by the people today, people use these applications to build a social relation and to stay connected with other user who shares the same interest, activity or real – life connection. Social Networking Sites have connected millions of people it is one the most popular internet activity today. Some of the popular Social Networking Sites are Facebook, Twitter, LinkedIn, Google Plus etc. As there are pros of these there are cons as well, as the people populate their online profile with a plethora of information that aims at offering a complete and correct information of themselves. Attackers use this information and form a fake account either in the same or different social networks and, therefore fool other users into forming trusting social relations with the faux profile. The Purpose of the paper is to identify the fake accounts and to compare different user activity algorithm in order to select the most appropriate. The proposed method address the problem by using two level of filtering, first level of filtering is done using Profile Matching using Fuzzy-sim algorithm, second level of filtering is User Activity Matching using three algorithms Predictive FP Growth Algorithm, Eclat, Apriori Algorithm and a comparative analysis is done to find most efficient algorithm among the FP growth, Eclat and Apriori.*

Keywords –online social network, social networking site, area of interest, social networking site 1, social networking site2

I. INTRODUCTION

Social networking has ended up as everyday interest on the internet these days, the recognition of online social networking sites is getting higher every day due to the friendliness delivered in the websites and technological development.[1] Facebook has around 500 million users and suppressed Google, LinkedIn hosts profiles for more than 70 million individuals and 1 million for businesses. As the majority of customers are not acquainted with privateer issues, they frequently disclose a huge amount of private information‘s on their profiles that can be accessed by all the users in the network. [3] In profile cloning, where a person other than the legitimate owner of a profile creates a new profile in the same or other social networking site wherein he copies the information by means of doing so, he creates a fake profile impersonating the valid owner and invite the victim contacts to form a social link and may exchange messages, they establish a social link and they will have complete access to the other users private information’s. Users might also keep profiles in more than one social networking site, their contacts, especially the more distinct users have no way of knowing if a profile encountered in a social networking site has been created by the genuine user or not.

A study [1] by Gross et al revealed that only 0.06%of the users hide visibility of information such as interests and relationship etc.. 99% twitter user had a default privacy setting. As online Social Networks consisting of Facebook and Google+ are getting progressively more embedded in people‘s daily lives, personal records turns into effortlessly uncovered and abused. Information harvesting, by the operator, malicious customers, and third party commercial business are being recognized as a great protection problem.

The purpose of this project is to overcome this sort of attacks, the system uses two levels of filtering for identification of fake profiles the method considers two Social Networking Sites SNS1 and SNS2, it matches the user1 profile in SNS1 with user1 profile in SNS2. The method adapted uses two levels of filtering, first level is User’s Profile matching and the second level is User’s Activity Matching. Fuzzy Sim algorithm is used for profile matching and for User Activity matching three different algorithms are used they are predictive FP Growth algorithm which is modified from the original FP growth Algorithm, Eclat, Apriori algorithm are used and a comparative analysis is done to find the best suitable algorithm among the three for User Activity matching.

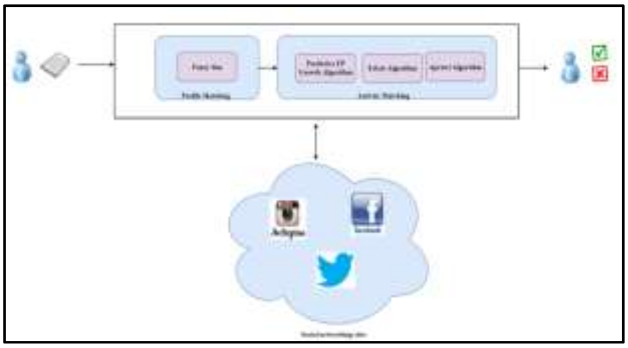


Fig-1: System Design.

II. LITERATURE SURVEY

People who intend to clone a profile could use different techniques able manual cloning automatic cloning using information from the victims homepage in addition information like profile photo first name last name extra are usually available publicly as default account setting this helps the attacker to have easy access to the user profile details, attacker use this details and create fake profiles. The most dangerous is automated profile cloning which could create many clone accounts and collect much more confidential information. The automated tool had to first break the CAPTCHA, it is an image containing hidden sequence of letters or digits, it's a computer program or system used to distinguish human from machine input. Unfortunately, many SNS allows many attempt to enter the captcha, and there many tools used to break them making it less secure and inefficient.

One of the method presented in [2] describes a solution for profile cloning in linkedIn the solution has three main components Information distiller, Profile Hunter, ProfileVerifier. Information distiller extract Key information from the current profile of the user, the profile should contain unique attributes for each specific user this information is used by the query to find the similar profile which may be a clone profile. Profile Hunter searches for the similar profile in different social networking sites finally profile verifier is presented with the suspicious profile details where similarity is calculated by comparing it with the predefined threshold, the fake profiles where identified . But the drawback of this method was it was applied only on one social network.

Other solution described in [4] using similarity comparisons the attributes were compared with fixed threshold, if result was less than the fix threshold then the profile were classified as clone profile else as genuine profile .Different methods were also used for similarity comparison such as attribute similarity based profile cloning detection, attribute similarity, dice's coefficient methods.

IP tracking methods were also employed the drawback was attacker could use the IP spoofing a common technique to change the IP address.

III. METHODOLOGY

The proposed system performs cross site profile cloning for identification of fake user, the profile of user1 in SNS1 is compared with User1 profile in SNS2, fuzzy- sim is used for Profile Matching, and three unsupervised techniques are applied for User Activity Matching namely Predictive FP growth, Eclat, Apriori .Predictive FP Growth is modified from the original FP Growth algorithm and finally comparative analysis is performed to find the best suitable algorithm for user activity matching. The method designed for this paper is discussed below, experiment was done using custom made social networking sites.

A. User activity matching
User Profile Matching (Fuzzy Sim)

Step 1: Scan the User (SNS1 and SNS2) Database and Extract the Constraints or Parameters.
Retrieve user profile details such as first name, last name, school, college, hometown, city etc.

Step 2: Calculate the Similarity..
Compare 2 user constraints, S (1,2)=No of Characters Matching/Total no of characters in Constraint .

Step 3: set threshold value to consider the constraints
If (value >0.5) Consider those constraints.

Step 4: Check if (constraints matching count>min_rec)
Set the number of constraints to match for fake account prediction - min_rec =8 constraints

◦ Genuine
else
◦ Fake

B. User Activity Matching

Predictive FP Growth

Step1: Retrieve shared information from the database (User1 — OSN1).
The information shared by the user is used for identifying the users area of interest.

Step2: Tokenization [keyword extraction method – removing the stop words and retrieving the keywords]
Get the user1 messages, remove the stop words, extract the keywords.

Step 3: Clustering the messages shared by the users (grouping of similar objects)
The keywords are compared with the predefined dataset (the set of keywords) the messages are clustered into different content type.

Step 4: Check if(matching count>min _ rec) [number of messages to compare]
A user must share minimum number of messages, for identification of his activity pattern.

Step 5: Retrieve shared information from the other database (User2 — OSN2).
Trace a user (OSN2) the area of interest, using the following steps

Step 6: Tokenization [keyword extraction method – removing the stop words and retrieving the keywords]

Step 7: Clustering the messages shared by the users.
By comparing with the predefined dataset (the set of keywords)

Step 8: Check if(matching count>min_rec)
A user must share minimum number of messages, for identification of his activity pattern.

Step 9: Compare the present user Area of interest (AOI) OSN1 with the previous user AOI OSN2.
If (AOI Matches)

- Genuine
  else
- Fake.

Eclat

STEP 1: Scan the dataset and form a transaction Id(tid) list
The result forms C1 (Candidate one item set) for each of the datasets.

STEP 2: Generate L1 (Frequent one item set).
L1 is generated from C1, compare candidate set item's support count with minimum support count if support _ count of candidate set items is less than min_support then remove those items the rest of them forms L1.

STEP 3: C2 (Combination of two items sets) is formed using L1.
Intersecting the tidlist of {a} with the tidlist of all the other item, which forms the tidlist as {a, b},{a, c},{a, d}….={a} conditional database.

STEP 4: Repeat the steps
To form C3(combination of three items sets), L3,C4,L4.. till C becomes null set, i.e. with only three item set present in L3, C4 cannot be formed C4 is combination of four item sets.

STEP 5: To form the Frequent item set (L).
L is a super set of L1, L2, L3. For each item in the frequent item set generate non-empty subsets.

STEP 6: To generate confidence

$$\text{Confidence }(X\text{->}Y) = \frac{\text{Number of records contain X}}{\text{Number of record contain both X and Y}}$$

for each non empty subset confidence is generated if the confidence generated is less than the minimum confidence then remove those item set, else consider those to form the pattern (Strong Association Rule)

Apriori

STEP 1: Scan the data set and determine the support(s) of each item.
C1 (candidate set) the table contains the support count of each item present in dataset.

STEP 2: Generate L1 (Frequent one item set).
Each candidate set item's support count is compared with minimum support count. If the support_count Of candidate set items is less than min_support then remove those items the rest forms L1

STEP 3: To generate C2.
C2 is generated using L1 called the join step, the condition for joining lk-1 is joined with lk-1 to generate set of candidate K

STEP 4: L2 is generated from C2
Compare C2 support count with minimum support count if less, then remove those item this gives the item set L2

STEP 5: To generate C3,L3,C4,L4..
Repeat the step until C becomes the null set, i.e., with only three item sets present in L3 ,C4 cannot be formed as C4 is combination of four items set.

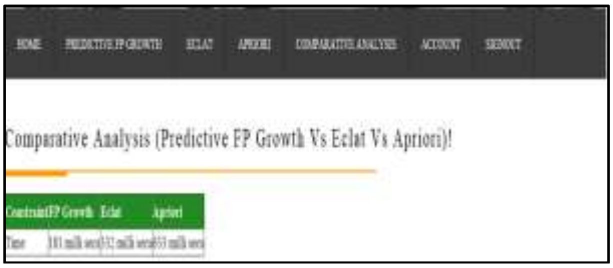
STEP 6: Generate Frequent item set(L).
Is generated using all L1,L2,L3..for each item in the frequent item set in L generate non-empty subsets.

STEP 7: Generate Confidence

$$\text{Confidence }(X\text{->}Y) = \frac{\text{Number of records contain X}}{\text{Number of record contain both X and Y}}$$

For all non empty subset confidence is generated if the confidence generated is less than the minimum confidence then remove those item set, else consider those to form the pattern (Strong AssociationRule).

IV. EXPERIMENTAL RESULTS AND SCREENSHOTS



Comparative Analysis (Predictive FP Growth Vs Eclat Vs Apriori)!

Contraint	FP Growth	Eclat	Apriori
Time	181 milli secs	132 milli secs	153 milli secs

Fig-2: Shows the execution time taken by each algorithm.

Predictive FP growth is the best suitable algorithm with least Execution time.

The time taken by Predictive FP Growth algorithm is 181 milliseconds, Eclat takes 332 milliseconds, Apriori takes 633 milliseconds. The results prove predictive FP Growth is the best suitable algorithm for User Activity Matching.

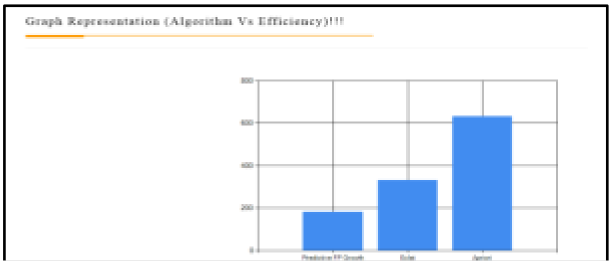


Fig- 3: Graphical representation of execution time.

The figure shows the graph plotted for the execution time taken for each algorithm.

V. CONCLUSIONS

The paper introduced the problem of colluding in Identity Clone Attack in OSNs. The solution applied has two levels of filtering, first an algorithm called Fuzzy-sim is used for the profile compassion. Second, algorithm called predictive FP Growth, Eclat, Apriori algorithms are used for pattern prediction of user activity. The Predictive FP Growth algorithm has few changes compared to the original FP growth Algorithm.

Comparative analysis results prove, the best algorithm suitable algorithm for User Activity matching is the modified predictive FP Growth algorithm, next best is Eclat algorithm and then the Apriori Algorithm, the predictive FP Growth Algorithm takes the lest execution time among the three algorithm, then Eclat and Apriori with takes the maximum time for execution. Finally, the graph show the quality of the output of classifier.

In future application could use some other algorithm which is more efficient than the algorithms used, and content classification can be applied on images and videos.

VI. REFERENCES

1. Comparing the Performance of Frequent Pattern Mining Algorithms.

Dr. Kanwal Garg Assistant Professor, Dept. of Computer Science and Applications, Kurukshetra University, Kurukshetra,Haryana,Deepak Kumar M.Tech Research Scholar, Dept.of Computer Science & Applications, Kurukshetra University, Kurukshetra, Haryana, India.

2. Detecting Social Network Profile Cloning.

Markatos Institute of Computer Science Foundation for Research and Technology Hellas {kondax, polakis, sotiris, Markatos Georgios Kontaxis, Iasonas Polakis, Sotiris Ioannidis and Evangelos P..

3. Friend Recommendation Framework for Social. Networking Sites using User‘s Online BehaviorMd. Mehedi Hasan1, Noor Hussain Shaon 2, Ahmed Al Marouf 3, Md. Kamrul Hasan4, Hasan Mahmud5, and Md.Mohiuddin Khan6 Systems and Software Lab (SSL), Department of Computer Science and Engineering (CSE) Islamic University of Technology (IUT), Gazipur, Bangladesh{1shovon10, 2shaon007, 3samcit41, 4hasank, 5hasan, 6mohiuddin}@iut-dhaka.ed

4. Preventing Colluding Identity Clone Attacks in Online Social Networks.

Georges A. Kamhoua1, Niki Pissinou1, S.S. Iyengar1, Jonathan Beltran1, Charles Kamhoua2, Brandon L Hernandez3, Laurent Njilla2, Alex Pissinou Makki4 1 School of Computing and Information Science, Florida International University, Miami, FL 33173 2 Air Force Research Lab, Cyber Assurance Branch 3 Computer Science, University of Texas at Rio Grande Valley, Brownsville, TX 78521 4Ransom Everglades School, Miami, FL 33133 {gkamh001, jbelt021, pissinou}@fiu.edu;{Charles.kamhoua.1,laurent.njilla}@us.af.mil;iyengar@cis.fiu.edu; Brandon.l.hernandez01@utrgv.edu, 7amakki@randomeverglades.org

5. Profile Cloning Detection in Social Networks.

Piotr Bródka, Mateusz SobasInstitute of Informatics Wrocaw University of Technology.