International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Email:editorijoer@gmail.com http://www.ijoer.in

Vol.4., Issue.2., 2016
(Mar-Apr)

RESEARCH ARTICLE

ISSN: 2321-7758

AN EFFICIENT WAY OF DETECTING CLONED PROFILES IN ONLINE SOCIAL NETWORK USING SIMILARITY MEASURE TECHNIQUES

TEDROS HAILEMARIAM, SUMAN TANWAR
Department of CS and IT, Symbiosis Institute of Technology, SIU
Pune, India

ABSTRACT

In the present day social network sites are growing fast and becoming one part of the social life for communicating and sharing information between communities and groups. It is also becoming a good environment for business organizations and companies to promote their products by creating a page or profile. Online Social Network sites enable a particular user to create and to set his/her profile by filling detail information like name, address, profile photo, job status etc… which makes one user unique from other users in a particular Online Social Network. It is obvious that Online Social Network sites are important on multiple sides however there are some problems raised related to privacy in which attackers or adversaries create a cloned or fake profile to illegally steal personal information, defame people, advert non genuine products and distribute spams which are the hot research areas in current days. This paper proposes an efficient way to detect a cloned profile by calculating the similarity measure of profiles using cosine similarity measure for exact string matching, Jaro–Winkler distance measure for an approximate (fuzzy) similarity measure of attributes and image processing technique of bit similarity for measuring the similarity of two user's profile photo. In this case when the value of the similarity measure techniques goes above the given threshold it will be detected as cloned profile.

Keywords – Profile cloning, Cosine similarity measure, Exact string matching, Jaro-winkler similarity, Image similarity.

©KY Publications

1. INTRODUCTION

Social networking is a web technology which builds a social network or social relations among people who share similar interests and activities. These sites consist of a representation of each user profile, and different services which allow individuals to design a public profile, create a list of users with whom to share ideas and information. Social network sites like Facebook and twitter are becoming increasingly popular nowadays Facebook is becoming one of the largest social media site in the world. As of the Facebook reviews the number of users of Facebook is more than 1.2 billion and it is growing at the rate of 3% per week [1] [10].

Social networks are useful for communicating and sharing information, events and idea to make the life style easy besides it is playing a major role on facilitating business by creating awareness to the people who is already on the network and it is helpful for politician to share their ideas and to create large followers.

Beyond those all facts there are some issues that make it social network need some researches to solve problems. One of the problems of these sites is security which is potentially risky for users of online social networks in which adversaries steal personal details to perform illegal actions. Adversaries can create a cloned profile in two ways the first one is same site cloning in which an attacker crates a profile in the same social network site in which the victim is already available for example if someone named James is in Facebook an adversary will also create another faked or cloned James in that social network by stealing the real profile's information. The second type of cloning is called cross site cloning in which an adversary collect personal detail of victim's from one social network or other web pages and crates another cloned fake profile in the second social network in which the victim have not still joined or registered on [1] [2] [9]. After having a cloned profile an adversary sends a friend request to victim's friend to collect additional sensitive information and to steal their personal information and creates cloned profile for them too. In the real world most of users simply confirm a user request even if it is already in their friend list and most of social network users confirm request without checking whether it is real profile or not [2].

Adversaries use user's detail information to perform illegal actions regarding to their needs some of them use it for financial matter because in current days there are many web technologies which needs user's information to perform financial tasks like banking and online shopping in this case smart hackers tries to hack those web technology using user's personal details. Other adversary also use it for defaming people they do not like, distributing spam, online bulling, promoting of non-genuine products.

To address these problems a number of researches have been conducted and to some extent these problem are getting solved. In current days privacy and security is becoming a major problem in the field of computer science because it is carried out by hackers who are intelligent and capable of guessing how those cloned profile detection tools work and they can prepare for deceiving these detection tools by getting aware of how they works. So adversaries have several possibilities to protect their fake profile from being detected [3]. For solving problems like this it is needed to have a good detection mechanism which eliminates such conditions.

2. RELATED WORKS

In current days Identity cloning attack or profile cloning attack is becoming a very hot issue which causes a financial destruction for many online social network users to overcome these problems a number of researches have been conducted on detecting cloned profiles in social networks. Kiruthiga .S Kola Sujatha.PKannan.A in [4] presented a tool which classifies the user profile information using naïve Bayes classification and clustered it using K- means clustering to easily manage each profile finally they have used cosine similarity to find the similarity between each profiles. FatemehSalehiRizi, Mohammad Reza Khayyambashi, and MortezaYousefiKharaji in [2] proposed a system to detect a cloned profile in social network using both attribute similarity measure and strength of relationship measure but here when they are measuring the attribute similarity they do not care whether the strings or attributes are approximately the same or not because attackers use an approximately the same word for example attackers change the name James to Jamec which will not be detected using exact string matching.

G.Kontaxis, I.Polakis, S. Ioannidis and P. Markatos in [5] proposed a tool for LinkedIn social network and their detection process includes three steps such us information Distiller in which information is extracted from real profile, in the second step they have used Profile hunter in which a suspicious profile is collected from the social network for further process and in the third step profile verifier calculates similarity measure on the value of information fields. If the similarity score is the same with the legitimate information the profile will be detected as cloned profile but here they use only exact string matching.

M.A Devmane, Dr.N.K.Rana in [1] proposed a system which detects a cloned profile in both same site

cloning and cross site cloning in which they have present three steps such us extracting information from the user's profile, after retrieving the legitimates information it searches for the user's profile from same site or from other social networking sites and finally it calculates the similarity index to identify the cloned profile. MortezaKharaji and FatemehRizi in [6] in this paper they have proposed a method to detect cloned profile in online social networks by calculating the strength of relationship between users.

3. PROPOSED SYSTEM

The proposed system uses mathematical techniques to efficiently detect cloned profiles by calculating the similarity measure between the victim's and all of the suspicious profiles of the social network. This approach consists of four steps to implement the detection of cloned profiles in the social network and in each techniques of the similarity measure there is a given threshold which is compared with the similarity measure of items of the two profiles to see the degree of similarity between of the two items of the profiles. The proposed system performs similarity measure for each item and decides whether it is cloned or not based on the given threshold.

Extraction of Legitimate attributes

Extraction of attributes is the first step of the cloned detection system in which the victim's attribute are retrieved from the online social network. In this stage all user's attribute are extracted from database which includes first name, last name, gender, country, location, education, email, profile photo etc. which uniquely identifies a particular user of the online social network. The extracted detail information of the user is used as a query for searching suspicious profiles and it performs similarity measure between victim's and suspicious profile in the next stage.

Attribute similarity(exact string matching)

In this stage an attribute similarity measure is calculated between the real profile and cloned profile by applying an exact string matching between of each attribute value of the profiles. By implementing cosine similarity measure the proposed system searches and retrieves all profiles from the entire online social network that have similar attribute value with the victim's attribute value. Since some adversaries are smart enough and they did not attempt to completely clone the victim's personal detail instead they leave some fields empty or replace it a particular attribute value of the real profile by some other value which does not match the victim's profile so as they can deceive and confuse others.

In the string matching if the value of items of two profiles is the same the particular attribute of the suspicious profile will be weighted as 1 otherwise it will be assigned to 0 which will be used for cosine similarity measure.

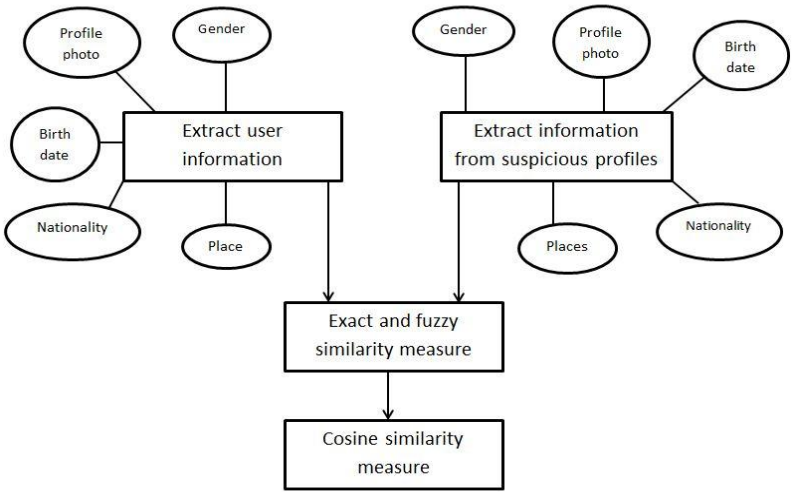

Fig1. Clone profile detection mechanism

Let Rpa and Spa be a real profile's attribute value and suspicious profile's attribute value respectively and w is a weight assigned to each item value i.

$$w_i = \begin{cases} 1, & Rpa_i = Spa_i \\ 0, & Rpa_i \neq Spa_i \end{cases}$$ ---------- (1) [2]

The weight of each attribute value of the real profile is assigned by default to 1 then using these weighted values it calculates the degree of similarity between the real and suspicious profile using cosine similarity measure which uses the weight of both suspicious and real profile's item value as a parameter.

$$Sim(A,B) = \frac{A.B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i{}^2}\sqrt{\sum_{i=1}^{n} B_i{}^2}}$$ ------ (2) [7]

When the result of the calculated cosine similarity measure goes above the limited threshold the suspicious profile detected as a cloned profile otherwise the system performs an approximate similarity measure between the attribute values of each item to increase the efficiency of detecting cloned attributes of the suspicious profile.

Approximate attribute similarity measure

An adversary who is aware of the above detection mechanism tries to deliberately make mistakes by misplacing and changing the sequence order of characters of an attribute value to deceive an exact string matching detection system. Since attribute values are short strings and adversaries try to change the sequence order of characters for example changing the first name James to Jamec deceives many online social network users though by doing such kind of activities they try to be a real users so to overcome and eliminate this problem the proposed system applies an approximate attribute similarity measure using Jaro–Winkler distance technique. If the Jaro–Winkler string similarity measure between the real value and the suspicious profile's value goes above the given threshold the particular suspicious attribute value will be detected as a cloned value and its weight will be 1.

$$d_j = \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3}\left(\frac{m}{s1} + \frac{m}{s1} + \frac{m-t}{m}\right) & otherwise \end{cases}$$ ----------

(3) [8]

Where m and t are the number of matching characters and half the number of transpositions respectively and dj is the jaro distance for string s1 and s2.

$$d_w = d_j + \left(lp(1 - d_j)\right)$$ ---------- (4) [8]

d_j is the Jaro distance for strings s1 and s2 and l represents the length of prefixes from the start of the word up to a maximum of four characters and p is a constant scaling factor and its standard value in Winkler's work is $p = 0.1$.

The above two formula of Jaro and Winkler calculates the degree of similarity between two given strings of attribute value the higher the Jaro-Winkler score indicates the higher the similarity degree between the strings. If the Jaro-Winkler's value equals to 0 it means there is no similarity between the strings and if the score equals to 1 it means there is an exact much between the two strings. The suspicious profile's attribute value will be detected as a cloned item if the scored Jaro-Winkler's value goes above the given boost threshold.

Profile photo similarity measure

Profile photo is an important attribute which most uniquely identifies social network users from each other for this reasons most of the adversary uses the victim's photo as their profile photo to deceive others and to look more legitimate. To overcome this problem image comparison will be conducted to find the similarity between two profile photo using image processing techniques. The image comparison will be takes place between all of the profile photos posted by the real profile and the suspicious profile if a similarity is found about any of the previously posted profile photos then it will be detected as a cloned item. Some smart adversaries change the format, size and other characteristics of the original photo to make a little difference to the original one and besides they will try to be not detected with exact image similarity and comparison techniques. Image processing techniques such as gray scale and image bit similarity will be used for comparing the similarity of two profile photos in the proposed system.

Image subtraction for measuring the similarity between two given profile photos.

$$R_{Photo} = imageread(realprofile)$$
$$S_{Photo} = imageread(suspiciousprofile)$$

$R_{Photo} = setimagesize(realprofile,)$
$S_{Photo} = setimagesize(suspiciousprofile)$
$R_{Photo} = change\_to\_RGBgray(realprofile)$
$S_{Photo} = change\_to\_RGBgray(suspiciousprofile)$
$Photo_{sim} = ImageBitDifference(R_{photo}, S_{photo})$ - (4)

A suspicious profile is detected as a cloned profile if it detected by all of the above similarity measurement techniques and if it goes beyond the given threshold t. To make the detection system more efficient the proposed system checks the joining date of both victim's and suspicious profile because in most cases adversaries create a cloned profile of an existing one so joining date will be used as one parameter to decide whether a given profile is cloned or not. After detecting the cloned profile the system sends notification to all friends of both profiles to let them know about the people claiming to be someone else.

An algorithm of the proposed system

Let S be a set of all user profiles and $RP, SP \in S$ where Rp and SP are real and suspicious profiles respectively and let i be a profile item of both RP and SP in which iR and iS represents attribute value of each item of both real profile and suspicious profile respectively. trepresents a threshold value of the similarity measure technique which will be used to decide a certain action.

1. Extract user's information i_R
2. For all i_R and $i_S \in i$ do
3. If $i_R = i_S$
4. Assign a weight of 1 to i_S, $W(i_S)=1$
5. Else if $d_w > t$ of $d_w(i_R, i_S)$ do
6. Assign a weight of 1 to i_S, $W(i_S)=1$
7. Else
8. Assign a weight of 0 to i_S $W(i_S)=0$
9. End if
10. End for
11. Set a weight of 1 for each i of R_P
12. $Sim(A,B) = Cos\theta(R_P, S_P)$
13. If $Sim(A,B) > t$ do
14. List out the detected profile
15. If $Photo_{sim} > t$ OR S_P(Join Date) > R_P(Join Date)
16. Block the detected profile S_P and send Notification to friends of both profiles
17. Else
18. Send notification to friends of both profiles
19. End if
20. End if

Fig.2 Algorithm of the proposed system

4. EXPERIMENTAL RESULTS

In this experiment we have collected real person's information to conduct the clone detection mechanism the credentials or attributes used for the detection process are First Name, Last Name, gender, Birth date, work, position, status, school and/or university name, city, nationality and profile photo. These attributes are used to search a suspicious profile from the social network. The detection approach validated using willing user's information who was interested to participate on the clone profile detection work. In this detection framework collecting of real user's information is performed and after having the original profile we add a cloned profile of the original profile deliberately because it is difficult to find real identities and their clone suspicious profiles on online social networks. The clone profile detection process compares the result of the proposed system and other methods of clone detection system.

Attribute similarity index

The similarity index shows the degree of similarity between two attribute of different profiles of the same item. In TABLE 1 below shows the cosine similarity index matrix using exact string matching which will be compared with TABLE 2 which includes both exact string matching and Jaro-Wnkler approximate similarity measure techniques.

The result of cosine similarity measure indicates the degree of similarity of two profiles of a given online social network. If the cosine similarity measure result ranges between 0.8 and 1 the suspicious profile will be considered as cloned profile which will be further confirmed by friends of both profiles of the given online social network and social network joining date comparison will be conducted to eliminate false positive and false negative problems. TABLE1 shows the similarity of two profiles only using exact string matching which is less efficient to detect cloned profiles and TABLE 2 shows the improved way of detecting clone profile using Jaro-Winkler fuzzy (approximate) string similarity.

Table 1 Result of cosine similarity using exact string matching

Cosine Similarity Index (matrix)	Profile 1	Profile 2	Profile 3	Profile 4	Profile 5	Profile 6
Profile 1	1	0.679	0.620	0.392	0.392	0.554
Profile 2	0.679	1	0.554	0.392	0.392	0.392
Profile 3	0.620	0.554	1	0.392	0.480	0.554
Profile 4	0.392	0.392	0.392	1	0.554	0.392
Profile 5	0.392	0.392	0.480	0.554	1	0.392
Profile 6	0.554	0.392	0.554	0.392	0.392	1

Table 2 Result of cosine similarity measure using exact string matching and Jaro-Winkler approximate (fuzzy) string matching.

Cosine Similarity Index (matrix)	Profile 1	Profile 2	Profile 3	Profile 4	Profile 5	Profile 6
Profile 1	1	0.877	0.733	0.620	0.554	0.733
Profile 2	0.877	1	0.620	0.554	0.480	0.679
Profile 3	0.733	0.620	1	0.392	0.554	0.554
Profile 4	0.620	0.554	0.392	1	0.620	0.679
Profile 5	0.554	0.480	0.554	0.620	1	0.554
Profile 6	0.773	0.679	0.554	0.679	0.554	1

The comparison between TABLE 1 and TABLE 2 is clear in which TABLE 2 uses the approximate similarity (Jaro-Winkler) algorithm to increase the efficiency of clone profile detection.

5. CONCLUSION

In current days fake cloning attack or fake identity attack is becoming the main problem of social networks in which a million of people are stolen their identity which cost them a billion of dollars [2]. This paper presented an efficient way of detecting fake cloned profiles in social networks to overcome the problem of identity theft. This approach uses three similarity measure techniques such as cosine similarity for attribute similarity, Jaro-Winkler for fuzzy string matching and image processing for profile photo similarity measure image subtraction and histogram measure. In this paper a profile is detected as cloned profile if the value of similarity measure of each technique is more than the given threshold.

REFERENCES

[1] M.A Devmane, Dr.N.K.Rana, 2014, "Detection and Prevention of Profile Cloning in Online Social Networks", IEEE international conference on recent advances and innovations in engineering, Jaipur India.

[2] Fatemeh S.R, Mohammad R. K, Morteza Y.K, 2014, "New Approach for Finding Cloned Profiles in Online Social Networks", ACEEE network security,

[3] Manuel.E, Gianluca.S, Christopher.K, Giovanni.V, "COMPA: Detecting Compromised Accounts on Social Networks".

[4] Kiruthiga. S, Kola Sujatha.P, Kannan.A, "Detecting cloning attack in social networks using classification and clustering techniques", 2014, IEEE international conference on recent trends in information technology.

[5] Georgios K, Iasonas P , Sotris I, Evangelos P. M, 2012, "Detecting social Network Profile Cloning", IEEE international workshop on security and social networking.

[6] Morteza Y.K, Fatemeh S.R, 2014, "An IAC Approach for Detecting Profile Cloning in Online Social Networks", Isfahan Iran.

[7] https://en.wikipedia.org/wiki/Cosine_similarity

[8] https://en.wikipedia.org/wiki/Jaro%E2%80%93Winkler_distance

[9] Mauro C, Radha P, Marco S, 2012 ," FakeBook: Detecting Fake Profiles in On-line Social Networks", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

[10] Leyla B, Thorsten S, Davide B, Engin K, 2009, "All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Network", ACM 978-1-60558-487-4/09/04

[11] P. Joshi and C. Jay Kuo, "Security and Privacy in Online social network: A survey", In Proceeding of IEEE International Conference on Multimedia and Expo, pp.1-6, 2011.

[12] Domingo Ferrer, A. Viejo, F. Sebé and Ú. González-Nicolás,"Privacy homomorphism for social networks with private relationships", Computer Networks Journal, Published by Elsevier, pp. 3007-3016, 2008.

[13] H. Gao , Jun Hu ,T. Huang , J. Wang and Y. Chen, "Security issue in online social networks", IEEE Internet Computing Journal , pp. 56-62, 2011.

[14] L. A. Cutillo, R. Molva and T. Strufe, "Safebook: A Privacy-Preserving Online Social Network Leveraging on RealLife Trust", IEEE Communication Magazine, pp. 94-101, 2009.

[15] http://www.news24.com/MyNews24/Facebook-cloning-How-its-done-and-how-to-prevent-it-20130530

[16] http://www.hoax-slayer.com/facebook-cloning-explained.shtml