International Journal on
Advanced Science
Engineering
Information Technology

Vol.11 (2021) No. 1
ISSN: 2088-5334

Clustered Approach for Clone Detection in Social Media

C. R. Liyanage[a,*], S. C. Premarathne[b]

[a] Department of ICT, Faculty of Technology, University of Ruhuna, Sri Lanka
[b] Department of IT, Faculty of IT, University of Moratuwa, Sri Lanka
Corresponding author: [*] ravihari@ictec.ruh.ac.lk

Abstract— **With the popularity of Online Social Networks (OSN), the number of different types of digital attacks has been increased. Identity Clone Attack (ICA) is one of the leading among them that illegally uses a genuine user's information by duplicating it in another fake profile. These attacks severely affect an identity since another malicious profile can misuse it. Hence, these clone profiles need to be identified and removed to increase the protection of users. This research introduces a novel approach to detect clone profiles on Facebook by using a clustering technique on its profile attributes and network connections. The detection process included three main stages: filter by name, cluster using weighted categorical attributes, and measure the strength of friend relationships among profiles, which follow one after another. Finally, the list of possible clones with their percentages representing the amount of duplicability to a given victim profile was presented as the model's output. With the Agglomerative hierarchical clustering algorithm and Jaccard similarity measurement, a low-average within-cluster distance in cluster density performance and a precision of 88.75% was shown in the results. Instead of suggesting the exact clones, the duplicability percentages make this approach more practical since there are many similar profiles but not clones. This methodology increases the model's adjustability to any other dataset as the selection of weights, thresholds, and clustering algorithm is done based on considering the distribution of the dataset.**

Keywords— **Clone profile; clustering; identity clone attack; network similarity; profile similarity.**

Manuscript received 13 Jul. 2019; revised 30 Aug. 2020; accepted 10 Dec. 2020. Date of publication 28 Feb. 2021.

I. INTRODUCTION

Recently Online Social Networks (OSNs) have become a significant part of people living where 2.46 billion of the global population is using it and is expected to reach around 2.95 billion in 2020 [1]. In addition to the many benefits facilitated by these networks, there are some protection and security risks caused by hidden identities called fake profiles. According to statistical estimations, Facebook has 81million fake accounts, whereas 5 percent of Twitter accounts are forged [2].

Identity theft or Identity Clone Attack (ICA) is one of the most popular attacks in OSN, and it is performed by profile cloning. Profile cloning is a way of stealing information from an existing user and creating new similar fake profiles using those details. Cloning a profile in OSN can be done with several intentions: to trick users, abuse financially, damage a person's reputation, and steal sensitive data of others [3].

When cloning profiles in OSN, the adversary first creates a fake profile using the victim profile's publicly available attribute information. A social network platform profile has a name, most probably a first and last name with another set of attributes such as birthday, hometown, and school to represent its identity. In profile cloning, most of the victim profile attributes will be copied by the clone profile. Usually, the name is the main feature both clone and the victim should have in common [4]. However, some of the attributes will not be copied by the same value; instead, some will be kept as empty or private. This is because clones can duplicate victim's features and maintain their privacy setting by making some of the attributes private. Also, an adversary can make some attributes public in which the corresponding victim had set to private such as birthday, where most of the users try to keep it private. According to the study [4], these activities may make the faked identity more realistic.

Typically, after cloning a profile, it will send friend requests to friends of the victim. Since the clone profile looks more like the genuine profile, friends of the victim will tend to accept a friend request from the clone without noticing that it is a duplicate profile of their friend [3], [4]. Hence, the adversary gets the chance to publish misleading content to the victim's friend audience using a clone profile to damage his good profile. Also, there can be some other problems caused due to the exposition of the victim's friends' private data to the adversary.

Before adding a friend to the network, a cautious user will first look for his friend list to check whether that user already exists or not. In that case, adding a considerable number of friends of the victim may not be manageable. Hence, the adversary tries to add the victim's recommended friends so that the clone becomes more genuine and makes it difficult for the victim to add those recommended friends [4]. The OSN platform usually generates the recommended friend list. They are the list of people who are not yet friends of the victim but having similar backgrounds or mutual friends between them.

As mentioned above, now a clone profile and the genuine profile will be very similar to public attribute values, friend networks, and recommended friend networks. Under these assumptions, the purpose of this study is to introduce a novel detection model that can use similarities in profile attributes and network details to find the possible clones for a given victim. The initial search space will be reduced in a more massive amount by filtering only the profiles with names like the victim's name to increase the efficiency. Next, these filtered profiles will go through clustering based on public attributes and filtering using network similarities. Finally, the suspect profile list will be presented with the amount of duplicability as a percentage. The model was developed for Facebook, which has the highest popularity, the largest number of user-profiles, and the highest number of fake profiles [1], [2].

The rest of the paper is organized as follows. Section II is arranged into four main parts; where the first part provides the details about the previous work done to resolve the problem of clone profile detection. The second and third parts explain the overview of the proposed approach and the dataset's information, respectively. Finally, the detailed model building approach was explained in the latter part of section II. Next, the evaluation of the results with the limitations and future work has been presented in section III, and finally, the overall conclusion has been given in Section IV.

II. Materials and Method

A. *Motivation from previous work*

The research area for detecting duplicate profiles in online social media networks has evolved recently, and most of the research findings were published after 2010. Since the research approaches are different depending on different OSNs, selecting the most suitable platform was the first most crucial step. Many researchers have addressed the problem of detecting identity clone attacks in a single platform, and the most common platform selections were Facebook [5]–[7], Twitter, Google+ [8], and LinkedIn. Moreover, some authors have used multiple platforms such as Google+ and Twitter and Facebook as their social environments [9], [10].

Different OSNs provide different types of quantitative and qualitative information such as name, gender, location, education, work, age, number of friends, comments, likes, and friend requests [2], [11]. However, due to different accessibilities for these data, researchers have used many techniques to gather them. The study [12] has used publicly available data, while in the study [7] the author has not used real data set for his implementation. Some researchers [6] have gathered data by creating fake experimental profiles called "Honey Profiles," and that was difficult than gathering data via online tools such as Facebook Graph API [9], [13], and Snapshot tool [6]. Finally, most of the past studies have gained datasets from external sources such as Barracuda Labs [14] and SNAP Library [9].

After gathering data, researchers have started their fake and clone detection process by experimenting with different techniques. The technique of modeling social graphs representing friend network connections was one of the commonly used approaches, and they identified duplicate profiles by analyzing friend patterns [3]. Another study [15] has used a social network of Facebook, and according to user similarities, it was divided into smaller communities. Inside these communities, the strength of the relationships was calculated. A case study [6] was performed to identify the fake nodes by considering network density, degree of nodes, and the correlation between nodes. Some algorithms like [16] have presented a method to detect clone profiles using a graph and network-based approach by analyzing the social network's structural similarity.

According to many researchers, the comparison of profiles based on calculated similarity measurements was a very effective clone detection technique. Some algorithms [17] have directly matched the strings in information fields to measure the similarities between profiles. The approach [18] has introduced a weighted dice similarity measurement to calculate the similarities and rank the selected attributes. An attribute similarity and friend network similarity approach has been discussed in some papers [4], [19]. They have considered three types of friend network features for analysis: friend list recommended friend list and excluded friend list.

Some algorithms have tried to solve this problem of identifying OSN fake clone profiles based on classification approaches. The study [7] proposes a three-step model to match two different profiles from different social media platforms. They have used a binary classifier for feature extraction based on users' information regarding friend requests and friend lists. Some approaches are there to find user profiles that belong to the same user over different social networks [12]. They have generated a similarity vector using a known dataset of paired accounts belongs to the same user across multiple networks. These vectors were then used as the training dataset for supervised classifiers such as KNN, Naïve Bayes, Decision trees, and SVM.

Finally, most of the OSN researchers were unable to validate their results on a real platform, while others [18] have performed result validation through social authentication in which general asking questions from the suspect clones about their profile friends' information. When these suspects are unable to answer the questions, they will be verified as clones. Another way of validation is asking for unique real-world ID from the suspects [15]. Furthermore, the researchers of some studies [20] have got the help of the Facebook security team to validate their findings.

B. *The proposed approach*

There are three main stages of the detection model. Not all the profiles will go through all these steps, but only the profiles filtered by each step will be forwarded to the next steps. First, the model will input the name of a victim profile who has claimed to find his clone profile/s. Under the assumption that the first step of making duplicate profiles is stealing another genuine identity, all the profiles with the same name as the victim will be filtered and forwarded. These selected set of profiles are referred to as the candidate profile set.

When a fake user wants to forge a genuine user, it is assumed that it will make the profile looks like that user. Hence, most of the public features will be the same in both profiles. Under these circumstances, the same name profiles, including the victim, will be sent to the second detection step, which is the clustering based on their weighted categorical attributes except the name. According to the cluster results, the candidate profiles grouped into the same cluster with the victim will be sent to the next step of detection as the suspect list. The method of selecting the clustering algorithm and the number of clusters will be discussed later in section D-3. By now, the filtered profiles have a higher similarity to the given victim based on their attribute features. Fig. 1 shows the steps of the detection stage.

The next step further verified the duplicability between the victim and each filtered suspect user profile by checking the combined friend and recommended friend network similarity. If this calculated profile similarity value between each victim and suspect is above a predefined threshold, then those profiles were selected as possible clone accounts of the victim, and the amount of duplicability of each suspect profile was given as a percentage. The way of defining the threshold value will be discussed later in section II-D-6 of the paper. For testing the accuracy of the model, an artificially generated profile set was used.

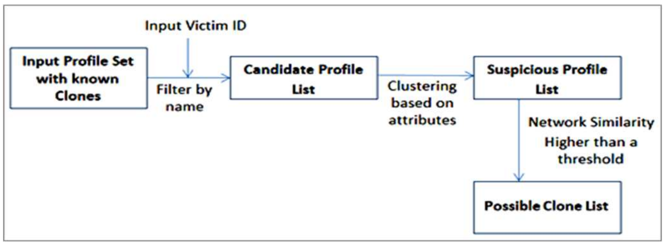

Input Victim ID
Input Profile Set with known Clones
Filter by name
Candidate Profile List
Clustering based on attributes
Suspicious Profile List
Network Similarity Higher than a threshold
Possible Clone List

Fig. 1 Detection steps of the proposed model

C. The Dataset

1) Data Collection: A Facebook dataset with many attributes and details of friend connections was downloaded from the online data repository of Stanford University (SNAP) [21] for the present study. The dataset has consisted of 4039 users, 88234 connections between them, and 26 profile attributes. A unique integer ID represented the network users, and an anonymous integer replaced all the attribute values. Only the subset of attributes given in Table 1 was used for further processing, and they were found by the literature [3], [11], [12], [15], [18], [22], [23] as frequently used attributes in detection methodologies.

TABLE I
ATTRIBUTES USED FROM PROFILES

No	Selected Attributes	No	Selected Attributes
1	First Name	6	Hometown
2	Last Name	7	Education School
3	Birthday	8	Work Employer
4	Gender	9	Work Position
5	Location	10	Work Location

2) Artificial Clone Set: Due to the difficulty of finding an originally verified clone profile set, this research study modified some of the existing profiles in the dataset as the clone set, which is to be 2% of the original dataset, and it was around 80 profiles. According to the characteristics stated in section I, clone profiles were given the same name as the victim, similar values for many attributes, and few NULL values.

It is known that a clone will not only duplicate a victim's attributes; rather, it will have similar network details due to the addition of the same set of users. Thus, the clones' friend networks were also modified to be similar (not exactly) to the victims' network. Furthermore, the dataset was created so that one genuine user can have one to three corresponding clone profiles.

3) Generate Recommended List: Recommended friend list for a victim was not randomly selected from their non-friend profile list. Instead, a set of non-friend users with a higher number of mutual friends were selected as recommended friends of a particular user when they have the same values on attributes such as hometown, location, school, and work employer [11], [24].

D. Model Building

1) Filter by Name: As stated in section I, the name is the key feature that will be the same in duplicate profiles. Hence as the first detection step, the users were filtered by their name, whose first name is like the given victim's name and forwarded to the next detection step. This step is essential to reduce the large population into a smaller one.

1) User Clustering - Attribute Weight Calculation: Before the second detection phase, which is the clustering, the attributes except the name were assigned with a weight according to the importance of them. Weights reflect the effect of each attribute during the process of detection and decision-making. Previous studies [18] have presented some formulas such as rank exponent, rank order centroid, and rank reciprocal. This study has calculated the weights using a method represented in the study [4], which considered the attribute value distribution in the dataset. For a particular attribute, the similarities between the values of each clone and victim pairs were calculated. Finally, the average of those similarities was taken as the weight of that attribute. In other words, this method of calculating the weights of attributes is more adaptive to any situation since this uses the known clone and victim pairs of any given dataset. Table 2 shows the process of estimating the weights briefly using an example.

TABLE II
ATTRIBUTE WEIGHT CALCULATION

User	Attr1	Similarity	Attr2	similarity
Victim1	315	0	763	1
Clone1	2103		763	
Victim2	410	0	103	1
Clone2	NULL		103	
Victim3	26	1	56	0
Clone3	26		89	

Weight/ Avg.	1/3=0.33	2/3=0.66

2) *User Clustering - Clustering Optimization:* In this study, the clustering technique was used to separate the profiles with similar attributes into the same groups and dissimilar attributes into different groups. The best number of clusters (K) considering the density performance for several clustering algorithms, namely kMeans, kMedoids and Agglomerative were calculated using the filtered candidate lists of each victim of the dataset. Then the average number of clusters for each clustering algorithm was found as in Table 3.

TABLE III
CLUSTERING ALGORITHMS WITH THE NUMBER OF CLUSTERS

Average Number of Clusters (K) with Density Performance		
KMeans	**KMedoids**	**Agglomerative**
7	6	6

Finally, due to the highest distribution performance shown as in Table 4, the Agglomerative clustering with complete Link Distance and corresponding K value was selected to cluster the profiles using nominal distance.

TABLE IV
CLUSTERING ALGORITHMS WITH THEIR DISTRIBUTION PERFORMANCES

Average Distribution Performances		
KMeans	**KMedoids**	**Agglomerative**
0.443	0.43	0.526

3) *Network Similarity Calculation-Friend and recommended friend network similarity:* Similarity is the measure of how much alike two data objects are. Profile similarity measurement is a value calculated to evaluate whether a given profile can become a clone of another account based on their networks. If the network similarity is higher than a predefined threshold value, then one of the two profiles is said to be cloned. This studied the friend network and recommended network information to calculate the network similarity among two profiles.

Based on the Jaccard similarity measurement, the following Equation 1 can be used to measure the similarity between friend lists (*F*) of two profiles. The Jaccard was selected among various similarity measurements due to its popularity and applicability in finding similarities of web-based applications and binary vectors [25].

$$\mathbf{S_{ff}} \quad (\mathbf{Pc, Pv}) \quad = \quad \frac{\mathbf{Fc} \ \cap \ \mathbf{Fv}}{\mathbf{Fc} \ \cup \ \mathbf{Fv}} \tag{1}$$

$\mathbf{S_{ff}}$ – The similarity between friend lists of two profiles.
Fc – Friend List of Clone Profile.
Fv – Friend List of Victim Profile.
$\mathbf{Fc} \cap \mathbf{Fv}$ – Common friends between Clone and Victim profiles (Mutual Friends).
$\mathbf{Fc} \cup \mathbf{Fv}$ – Total friends are available in the Clone and Victim profiles.

Based on the Jaccard similarity, the following Equation 2 can be used to measure the similarity between the friend list (*F*) of the clone profile and the recommended friend list (*RF*) of the victim profile.

$$\mathbf{S_{rf}} \quad (\mathbf{Pc, Pv}) \quad = \quad \frac{\mathbf{Fc} \ \cap \ \mathbf{RFv}}{\mathbf{Fc} \ \cup \ \mathbf{RFv}} \tag{2}$$

$\boldsymbol{S_{rf}}$ – The similarity between the friend list of the clone and recommended friend list of the victim.
$\boldsymbol{Fc}$ – Friend List of Clone Profile.
$\boldsymbol{RFv}$ – Recommended Friend List of Victim Profile.
$\boldsymbol{Fc} \cap \boldsymbol{RFv}$ – Common friends between two networks.
$\boldsymbol{Fc} \cup \boldsymbol{Fv}$ – Total friends available in two networks.

4) *Network Similarity Calculation - Aggregate Network Similarity:* Using Equation 3 the overall network similarity can be calculated by aggregating the network similarities between friend networks of victim and clone (*S_ff*) and the recommended friend list victim and friend list of clones (*S_rf*).

$$\mathbf{S_n} \quad (\mathbf{Pc, Pv}) \quad = \quad (\boldsymbol{\alpha}\mathbf{S_{ff}} \quad + \quad \boldsymbol{\beta}\mathbf{S_{rf}}) \tag{3}$$

$\boldsymbol{S_{ff}}$ – The similarity between friend lists of two profiles.
$\boldsymbol{S_{rf}}$ – The similarity between the friend list of the clone and recommended friend list of the victim.
$\boldsymbol{S_n} \ (\boldsymbol{Pc, Pv})$ – Aggregate Network Similarity between two profiles.

The importance of those two network similarities is different [4] where the importance of *S_ff* is higher than *S_rf* of the overall aggregate network similarity (*S_n*). Thus, α > β and they were calculated as α=0.9 and β=0.1 by taking the average *S_ff* and *S_rf* between all known clone victim pairs.

5) *Network Similarity Calculation - Similarity Threshold Generation:* Threshold similarity value was estimated by considering the aggregate network similarity values between the known clones and victim pairs. Instead of taking an average aggregate network similarity, the minimum among all known pairs was taken as the threshold to avoid losing some of the actual clone profiles without being detected by the threshold. An example was given in below Table 5.

TABLE V
SIMILARITY THRESHOLD CALCULATION

User	Aggregate Network Similarity	Similarity Threshold
Victim1 Clone1	0.85	0.8
Victim2 Clone2	0.8	
Victim3 Clone3	0.9	

6) *Network Similarity Calculation - Decision Making:* The minimum threshold taken by the known pairs was 0.93. All the suspect profiles with a similarity value higher than this value were given a percentage indicating how similar they are to the genuine victim. The application should have a real-time validation mechanism to verify the actual clone profiles.

III. RESULTS AND DISCUSSION

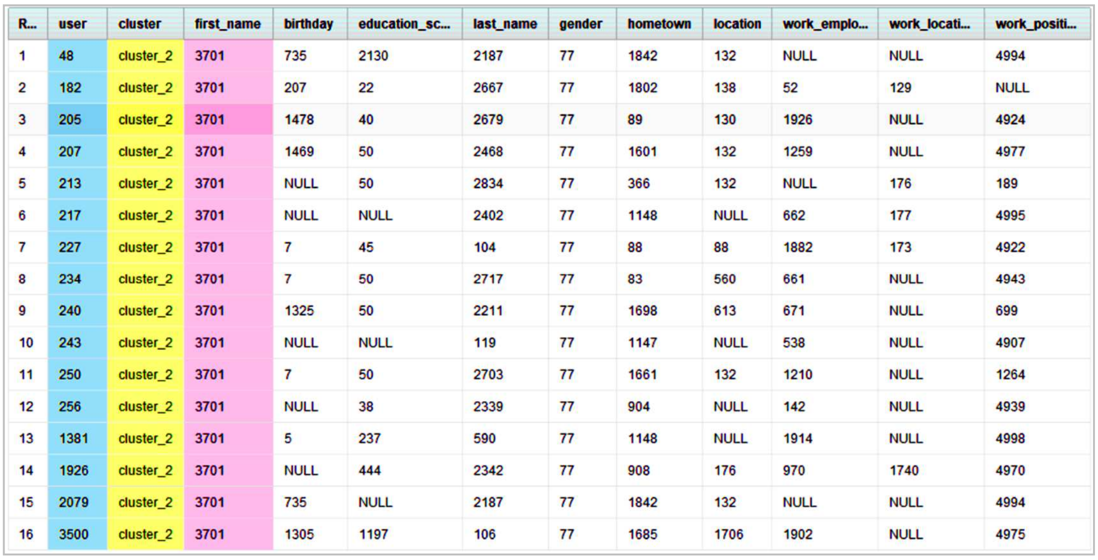
This research study refers to user filtration based on clustering and statistical similarity method. However, the clustering technique was a fully unsupervised learning mechanism; some known labels such as the clones and their victims were used in the result evaluation. This technique was used to evaluate how well the clustering matches the gold standard[26] classes of victims and clones. Nevertheless, this gold standard calculates some statistical values such as similarity threshold, attribute weights, and network weights.

A. Testing and Results

After building and training, the detection model was tested on the unknown dataset. The testing dataset was generated artificially with 3000 profiles using a data generating tool. The following figures show the corresponding outputs of a given victim with ID=48 and clone with ID=2079. From 3000 profiles, 94 were filtered (Table 6) by the name of the victim. Among those 94 profiles, only 15 profiles have been grouped with the victim (Fig. 2) from clustering. Then from these 15, only 4 profiles have been selected as the possible clones (Fig. 3).

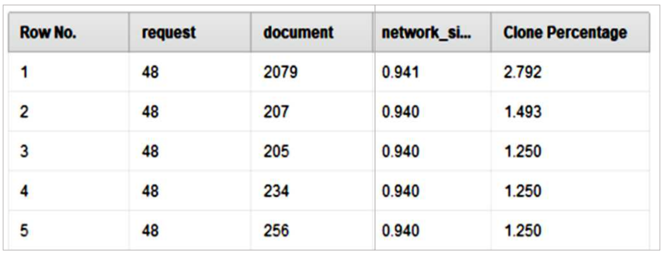
TABLE VI
NUMBER OF PROFILES PER CLUSTER

Cluster Model	
Cluster 0:	20 items
Cluster 1:	31 items
Cluster 2:	16 items
Cluster 3:	19 items
Cluster 4:	2 items
Cluster 5:	6 items
Total number of items:	94



R...	user	cluster	first_name	birthday	education_sc...	last_name	gender	hometown	location	work_emplo...	work_locati...	work_positi...
1	48	cluster_2	3701	735	2130	2187	77	1842	132	NULL	NULL	4994
2	182	cluster_2	3701	207	22	2667	77	1802	138	52	129	NULL
3	205	cluster_2	3701	1478	40	2679	77	89	130	1926	NULL	4924
4	207	cluster_2	3701	1469	50	2468	77	1601	132	1259	NULL	4977
5	213	cluster_2	3701	NULL	50	2834	77	366	132	NULL	176	189
6	217	cluster_2	3701	NULL	NULL	2402	77	1148	NULL	662	177	4995
7	227	cluster_2	3701	7	45	104	77	88	88	1882	173	4922
8	234	cluster_2	3701	7	50	2717	77	83	560	661	NULL	4943
9	240	cluster_2	3701	1325	50	2211	77	1698	613	671	NULL	699
10	243	cluster_2	3701	NULL	NULL	119	77	1147	NULL	538	NULL	4907
11	250	cluster_2	3701	7	50	2703	77	1661	132	1210	NULL	1264
12	256	cluster_2	3701	NULL	38	2339	77	904	NULL	142	NULL	4939
13	1381	cluster_2	3701	5	237	590	77	1148	NULL	1914	NULL	4998
14	1926	cluster_2	3701	NULL	444	2342	77	908	176	970	1740	4970
15	2079	cluster_2	3701	735	NULL	2187	77	1842	132	NULL	NULL	4994
16	3500	cluster_2	3701	1305	1197	106	77	1685	1706	1902	NULL	4975

Fig. 2 Profiles filtered to cluster 2



Row No.	request	document	network_si...	Clone Percentage
1	48	2079	0.941	2.792
2	48	207	0.940	1.493
3	48	205	0.940	1.250
4	48	234	0.940	1.250
5	48	256	0.940	1.250

Fig. 3 Final suspect clone list for victim=48

B. Model Performance

The following Fig. 4 was taken from the above victim-clone pair example, and the differences between the similarity percentage of the actual clone and the other predicted clones are distinguishable. Nevertheless, considering all the victim-clone pair examples the precision [TP/ (TP + FP), where TP – examples selected the actual clone as the possible clones with the highest clone percentage and FP – False Positive, examples did not select the actual clone as the possible clones with the highest clone percentage] was 88.75% and it was considered as a good performance.

Moreover, the system's performance depends on the clustering technique in which most of the similar profiles will be filtered out from a large sample. Hence the selection of a suitable clustering algorithm and a similarity or distance measurement is crucial. The density-based cluster performance evaluation was used to evaluate the clustering method's performance, and it gave relatively low average within-cluster distance values for most of the examples where it was -50.446 for the above example.

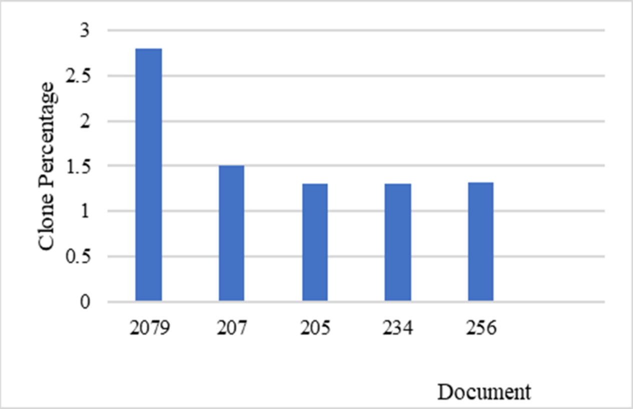


Fig. 4 Clone percentage distribution of possible clone list

C. Limitations and Challenges

Most previous studies conducted the experiments on OSN deal with getting a realistic dataset with all the required information. Here, due to the modification of the dataset by adding artificial clones and recommended friend lists, the original dataset's natural patterns were changed. Due to anonymous names replaced by integers, similar names with little changes will not be detected as similar. Further, the inability to validate the model on a real OSN platform by verifying the resulted clone profiles was another main limitation of the work.

D. Future Improvements

As future improvements to this work, the model can be tested with more than ten attributes to identify the relationships between different attributes and clone profiles. Further cross-platform detections where the clones are in a different platform, using string matching algorithms to match the actual text of a name when non anonymized features are given, can improve this work. Also, this research study's main future interest is to build a model to detect the actual person behind this clone who created the clone profile by analyzing the behavioral patterns of profiles in OSN.

IV. Conclusion

Due to the popularity of the platform and the simplicity of making profiles, the threat of creating clone profiles has been increased on Facebook. With this attack, users' personal information can be misused and can cause damages to their good reputation. This paper introduces a model with three primary stages to detect these clone profiles on Facebook, wherein at each stage, the amount of computation to be done was reduced by filtering profiles in each of the stages. This method was a simple but more effective method that also showed a higher precision. Furthermore, as most of the calculations are done considering the dataset's distribution, this model can be easily adjusted to a different dataset by only finding values for few parameters.

References

[1] Statista, "Social Media Statistics & Facts," 2017. [Online]. Available: https://www.statista.com/topics/1164/social-networks/. [Accessed: 30-Oct-2017].

[2] WordStream, "40 Essential Social Media Marketing Statistics for 2017," 2017. [Online]. Available: http://www.wordstream.com/blog/ws/2017/01/05/social-media-marketing-statistics. [Accessed: 10-Nov-2017].

[3] F. Rizi, M. Khayyambashi, and M. Kharaji, "A New Approach for Finding Cloned Profiles in Online Social Networks," Int. J. Netw. Secur., vol. 6, no. April, pp. 25–37, 2014.

[4] L. Jin, H. Takabi, and J. B. D. Joshi, "Towards active detection of identity clone attacks on online social networks," Proc. first ACM Conf. Data Appl. Secur. Priv. - CODASPY '11, p. 27, 2011.

[5] P. Dewan, S. Bagroy, and P. Kumaraguru, "Hiding in Plain Sight: Characterizing and Detecting Malicious Facebook Pages," pp. 193–196, 2016.

[6] K. Krombholz, D. Merkl, and E. Weippl, "Fake identities in social media: A case study on the sustainability of the Facebook business model," J. Serv. Sci. Res., vol. 4, no. 2, pp. 175–212, 2012.

[7] G. A. Kamhoua et al., "Preventing Colluding Identity Clone Attacks in Online Social Networks," in 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), 2017, pp. 187–192.

[8] M. A. Devmane and N. K. Rana, "Detection and prevention of profile cloning in online social networks," Int. Conf. Recent Adv. Innov. Eng. ICRAIE 2014, pp. 9–13, 2014.

[9] M. Torky, A. Meligy, and H. Ibrahim, "Recognizing fake identities in online social networks based on a finite automaton approach," 2016 12th Int. Comput. Eng. Conf. ICENCO 2016 Boundless Smart Soc., pp. 1–7, 2017.

[10] M. Egele, C. Kruegel, and G. Vigna, "COMPA : Detecting Compromised Accounts on Social Networks."

[11] P. Bródka, M. Sobas, and H. Johnson, "Profile cloning detection in social networks," Proc. - 2014 Eur. Netw. Intell. Conf. ENIC 2014, pp. 63–68, 2014.

[12] A. Malhotra, L. Totti, W. Meira, P. Kumaraguru, and V. Almeida, "Studying user footprints in different online social networks," Proc. 2012 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2012, pp. 1065–1070, 2013.

[13] R. N. Reddy and N. Kumar, "Automatic detection of fake profiles in online social networks," 2012.

[14] N. Kumar and R. N. Reddy, "Automatic Detection of Fake Profiles in Online Social Networks," National Institute of Technology Rourkela Rourkela-769 008, Orissa, India, 2012.

[15] M. Kharaji and F. Rizi, "An IAC Approach for Detecting Profile Cloning in Online Social Networks," Int. J. Netw. Secur. Its Appl., vol. 6, no. 1, pp. 75–90, 2014.

[16] M. Zabielski, R. Kasprzyk, Z. Tarapata, and K. Szkółka, "Methods of Profile Cloning Detection in Online Social Networks," MATEC Web Conf., vol. 76, 2016.

[17] G. Kontaxis, I. Polakis, S. Ioannidis, and E. P. Markatos, "Detecting social network profile cloning," 2011 IEEE Int. Conf. Pervasive Comput. Commun. Work. PERCOM Work. 2011, pp. 295–300, 2011.

[18] M. R. Khayyambashi and F. S. Rizi, "An approach for detecting profile cloning in online social networks," 2013 7th Intenational Conf. E-Commerce Dev. Ctries. With Focus e-Security, ECDC 2013, pp. 1–12, 2013.

[19] F. S. Rizi and M. R. Khayyambashi, "Profile Cloning in Online Social Networks," Int. J. Comput. Sci. Inf. Secur., vol. 11, no. 8, pp. 82–86, 2013.

[20] Q. Cao, X. Yang, J. Yu, and C. Palow, "Uncovering Large Groups of Active Malicious Accounts in Online Social Networks," Proc. 2014 ACM SIGSAC Conf. Comput. Commun. Secur. - CCS '14, pp. 477–488, 2014.

[21] J. Lescovec, "Stanford University - Data Repository," 2012. [Online]. Available: https://snap.stanford.edu/data/egonets-Facebook.html. [Accessed: 10-May-2018].

[22] S. Mazhari, S. M. Fakhrahmad, and H. Sadeghbeygi, "A user-profile-based friendship recommendation solution in social networks," J. Inf. Sci., vol. 41, no. 3, pp. 284–295, 2015.

[23] D. Dave, N. Mishra, and S. Sharma, "Detection Techniques of Clone Attack on Online Social Networks: Survey and Analysis," pp. 179–186.

[24] Facebook, "Finding Friends and people you may know," 2018. [Online]. Available: www.facebook.com/help/www/336320879782850. [Accessed: 10-Dec-2018].

[25] V. A and R. I. M. Dunbar, "Evolutionary Dynamics in Twitter Ego Networks," 2015. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/jaccard-coefficient. [Accessed: 23-Sep-2018].

[26] "Evaluation of clustering - Stanford NLP Group," 2009. [Online]. Available:https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-clustering-1.html. [Accessed: 03-Jan-2019].