Profile Cloning Detection in Online Social Networks

M. ZABIELSKI, Z. TARAPATA, R. KASPRZYK, K. SZKÓŁKA

mzabielski@wat.edu.pl, ztarapata@wat.edu.pl, rkasprzyk@wat.edu.pl, kszkolka@wat.edu.pl

Military University of Technology, Faculty of Cybernetics
Institute of Computer and Information Systems
Kaliskiego Str. 2, 00-908 Warsaw, Poland

Due to the emergence of online social networks, the importance of privacy on the Internet has vitally increased. Thus, it is important to develop mechanisms that will prevent our hidden personal data from unauthorized access and use. In this paper an attempt was made to present a concept of profile cloning detection in Online Social Networks (OSN) using Graph and Network Theory. Compering values of attributes of users' personal profiles and analysing structural similarity of networks, we identify attackers which steal users' identity.

Keywords: profile cloning detection, online social networks, violations of privacy.

1. Introduction

One of the aspect of worldwide accessibility to the Internet is social media which includes blogs, forums, photo-sharing platforms, social gaming, chat apps and social networks. According to the Statista portal [19] in 2018 there will be around 2,55 billion social network users in total and this is around a third of Earth's entire population (Figure 1).

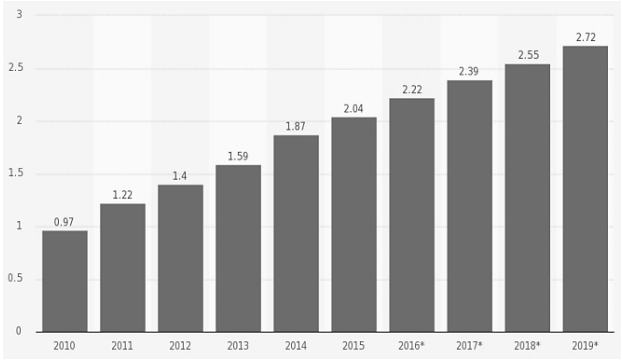

Fig. 1. Number of social network users worldwide from 2010 to 2019;
Source: www.statista.com

It is commonly known that online social networking is popular and it plays important role in people daily life. On the other hand, social networks users are mostly unaware of many threats existing in social networks. These threats can be divided into four categories: *classic threats* (malware, phishing attacks, cross-site scripting, etc.), *modern threats* (clickjacking, socialbots, **profile cloning attacks**, etc.), *combination threats* and *threats targeting children* (online predators, risky behaviour, cyberbullying). In this paper authors will focus only on profile cloning detection. The reader may refer to [18] for further details about current threats in social networks.

An user profile in the Online Social Network (OSN) can be represented as a set of features and links between other profiles that describe the person in that network. Depending on the type of OSN, a set of attributes and relationships that make up a profile may be different [2]. This proves there isn't an unique user social network profile on the Internet today, which gives the possibility to clone it. The method of *cloning an user profile* in OSN involves impersonating a victim by creating the best possible copies of their social profile within the same or different social network (*local or global profile cloning*) (Figure 2). In the simplest scenario, it includes building relationships and sharing profile data in the same manner as in the victim profile. As a result, an attacker is able to steal the identity of the victim in the same social network or different one.

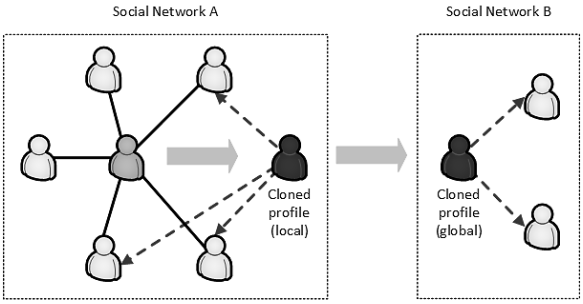

Fig. 2. Local and global profile cloning attacks

Nowadays there are some methods that detects a profile cloning attacks [3], [5]. Most of them utilizing Machine Learning algorithms [6], [7]. This, in most cases, results in analysing users' profiles in online social network comparing values of their attributes without taking into account relationships between users' profiles.

Another way of thinking about profile cloning pays attention that it is a process perform by hackers to conduct this kind of cyberattack. Having a knowledge about how hackers works gives an opportunity to detect symptoms of cyberattacks, which eventually implicates developing solutions to counteract or neutralize consequences. This approach is well described in [8].

We have developed the method to detect profile cloning which takes into account not only values of user profile attributes, but also relationships (links) between users' profiles in Online Social Networks. The paper extends the paper [17]. In particular an explanatory case study was described.

The paper is organized as follows. Section 2 contains proposition for revealing hidden attributes of social profile. In section 3 authors describes node similarity measure. Section 4 contains description of graph structural similarity. In section 5 the concept of profile cloning detection using methods described in sections 2, 3 and 4 is presented. Section 6 contains a simple case study.

2. Revealing hidden attribute values of social profile

For detecting a profile cloning it is important to have a complete information about user personal profile in Online Social Network.

Most often the reason of hidden attribute value are mechanisms, provided by Online Social Networks, which give us an opportunity to show values of attributes only for user-defined group of people in network. In most cases this approach is implemented by Access Control Lists (ACL). Despite their usability in protecting privacy aspect [1], [4], this situation impede our analysis in sense of profile cloning detection.

To provide a situation, when we know all about attribute values for each personal profile in particular Online Social Network, we provide a method, based on network model and k-nearest neighbours' algorithm, which allows us to estimate hidden values of attributes for appropriate personal profile. A general procedure for this shows Figure 3.

First of all, we have to provide analysing network, where nodes represents persons and edges describe a relation between two people. Moreover, a set of functions on nodes, which represent attributes of social profile and one function on edge, which defines force of relation has to be defined. In that network we choose a node, for which we want to reveal hidden attribute values – we denote it as w_s.

Next, we get the nearest neighbours of node w_s – that is a set of direct friends of considered person – and calculate for them a similarity measure between person w_s and his neighbors. The similarity function can be any function that gives a higher value for nodes that are similar in sense of attribute values. In our approach, we use function described by equation:

$$s(w_s, w) = d(w_s, w) * \sum_{a_i \in A} p_{a_i}(w_s, w) \qquad (1)$$

where:

$d(w_s, w) \in (0,1]$ – force of relation between person w_s and w. It can be measured, for example, as a number of messages sent to person w by w_s divided by numbers of whole messages sent in OSN by node w_s,

$$p_{a_i}(w_s, w) = \begin{cases} 1 \; if \;\; a_i(w) = a_i(w_s) \wedge a_i(w) \; known \\ 0 \; if \; a_i(w) \neq a_i(w_s) \vee a_i(w) \; unknown \end{cases} \qquad (2)$$

$a_i(w)$ – value of the *i*-th attribute of personal profile of person *w*;
A – set of personal profile attributes.

When a similarity is calculated for every neighbour, we choose *k* nodes (the best nodes) with the highest similarity measure to consider their attributes values. The parameter *k* gives us confidence that calculated value of hidden attributes for node w_s will not be dominated by only one node in Online Social Network. After choosing appropriate nodes, we calculate a frequency of appearing every value from the best nodes that are connected with hidden attributes. The most frequent value is taking as an answer. If we have more options to choose a value, we take a value with better similarity measure. We can repeat this procedure for every attribute that value is hidden.

The whole algorithm of revealing hidden values of attribute can be used for every node in a network, giving us an opportunity to estimate value for every attribute in every personal profile in OSN.

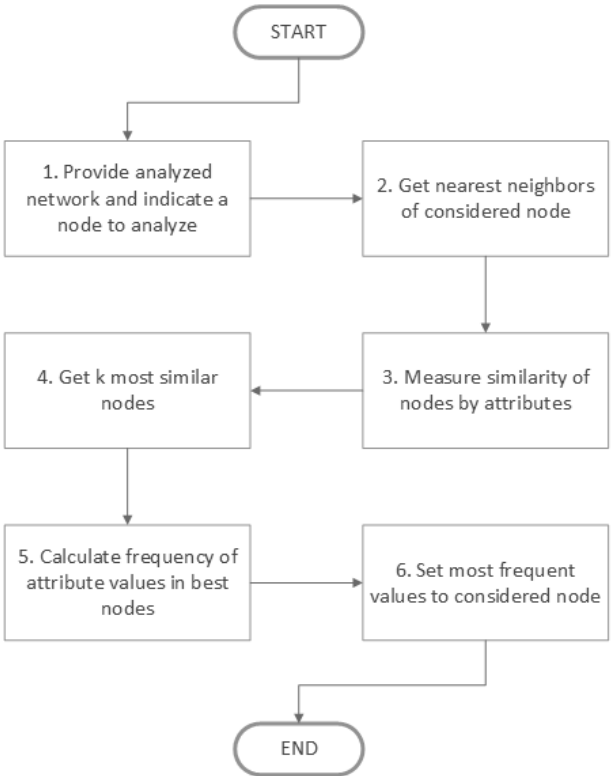

Fig. 3. Revealing hidden attributes of social profile procedure

There are some important issues that are worth discussing for a better use of this model. Firstly, because this method is partially derived from *k*-neighbours' algorithm, it is good to choose a parameter *k* that is odd. This approach ensures that chance to get a value using only a frequency of appearing value in best nodes is high, which gives us better estimation of attribute hidden value. From our experiments, it is also good to choose parameter *k* from range $[3, 9]$, depend on nearest neighbour amount of considered node. The more neighbours' person has the higher value of parameter *k* should be. Secondly, it is good to know which kind of attribute values we can get. A similarity measure defined in (1) is good for discrete values, but it can give worst results for continuous values. If we have knowledge about a value range for attributes, we can use mix of similarity measures, which probably gives us a better estimation for continuous values of attributes.

3. Node similarity

After revealing the most probable values of hidden attributes in user social profile, we have to provide a measure that helps us to detect an appropriate person from pattern network in our analysed network. To achieve this, we have to use a function which fulfils following conditions:

- it allows to simply compare attribute values between persons;
- it should be easy to interpret, which means that by using them we will be able to decide which personal profile, in sense of attribute values, is most similar to considered person's profile without any additional transformation of measuring function values;
- it should use a characteristic which are easy to obtain from network model.

According to the presented assumptions, for our purposes, we use a nodes' similarity function, which is described by (3):

$$ID(w_s, w) = \sum_{i=1}^{m} I_i p_{a_i}(w_s, w) \in [0,1] \qquad (3)$$

where:

m – number of attributes in personal profile;
$I_i \in [0,1]$ – the importance of the *i*-th attribute $\sum_{j=1}^{m} I_j = 1$;
$p_{a_i}(w_s, w)$ – comparing function defined by (2).

Function $ID(w_s, w)$ meet all the conditions defined for node similarity measure function. By using only simple comparing of attribute values, it is reduced to compare vectors of attributes that creates user personal profile. Interpretation of (3) is easy, because $ID(w_s, w)$ can take values from range [0,1]. Finally, using only attribute value and significance, we provide a method, which base only on characteristics described directly on network.

Having a measure of similarity between nodes we can obtain a global measure of node similarity between two networks (pattern and analysed networks) S and S', which gives us a better understanding of similarity of whole network. To do this we can formulate and solve (using for example Hungarian algorithm) optimal assignment problem to find the best allocation matrix $X = [x_{ij}]_{|W_G| \times |W_{G'}|}$ of nodes from network S to nodes from network S':

$$G_{ID} = \sum_{i=1}^{|W_G|} \sum_{j=1}^{|W_{G'}|} ID(i,j) \cdot x_{ij} \to max \qquad (4)$$

with constraints:

$$\sum_{i=1}^{|W_G|} x_{ij} \leq 1, j = \overline{1, |W_{G'}|} \qquad (5)$$

$$\sum_{j=1}^{|W_{G'}|} x_{ij} \leq 1, i = \overline{1, |W_G|} \qquad (6)$$

$$\bigwedge_{i \in \{1,\ldots,|W_G|\}} \bigwedge_{j \in \{1,\ldots,|W_{G'}|\}} x_{ij} \in \{0,1\}$$

where:

W_G – set of nodes in pattern network *S*;
$W_{G'}$ – set of nodes in analysed network *S'*.
Thanks that, we will be able to assess a "percentage" of similarity between considered networks and their persons.

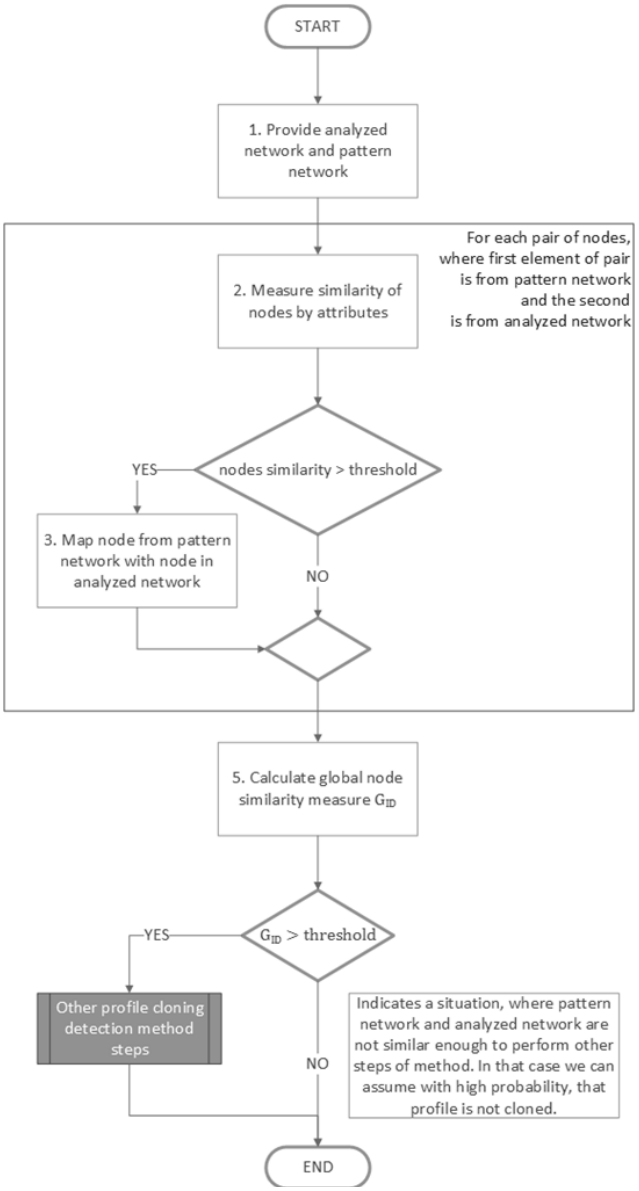

Fig. 4. Algorithm of mapping persons from pattern network to analysed network in profile cloning detection method

It is important to remember that solutions provided in (3) and (4) are only some kind of proposals how to calculate a similarity at the vertices and network level, which we use in our specific implementation. For profile cloning detection method, which will be described in detail in chapter 6, we can define any other methods, which fulfil the conditions described earlier. Some of them are defined in details in [9]. In general, a procedure which is used in our profile cloning method is illustrated in Figure 4.

4. Graph structural similarity

There are a lot of measures that we can use to assess similarity of graphs nowadays. In this section, we present most popular of it. Other methods and models presented in this paper are independent of the algorithm that is used for measuring graph similarity.

Graph isomorphism [12], [13], maximum common subgraph [12], minimum common supergraph [12], edit distance [12], [13], topological measures [12], [13] and iterative methods [10], [12], [13] are examples of structural graph similarity measures.

For the purpose of this research we propose graph structural similarity measure based on node neighbour matching. It is classified as one of iterative measures. The idea of it is as follows: two nodes are considered similar if they neighbours are similar [8], [9], [11], [14].
Let structural similarity measure over the nodes of two networks S and S' be represented as [9], [13]:

$$P_{node}(S,S') = \left[s_{ij}\right]_{|W_G|\times|W_{G'}|} = \lim_{k\to+\infty} Z_{2k} \tag{7}$$

where:

$$Z_{k+1} = \frac{BZ_kA^T + A^TZ_kB}{\|BZ_kA^T + A^TZ_kB\|}, k \geq 0,$$

A and *B* – transition matrices of *S* and *S'*, Z_0=**1** (matrix with all elements equal 1); A^T – matrix *A* transposition; $\|y\| = \sqrt{\sum_{i=1}^{|W_G|}\sum_{j=1}^{|W_{G'}|} y_{ij}^2}$.
Element z_{ij} of the matrix *Z* describes similarity score between the *i*-th node of the *S'* and the *j*-th node of the *S*. The greater value of z_{ij} the greater similarity between the *i*-th node of the *S'* and the *j*-th node of the *S*.

Having matrix $P_{node}(S,S')$, optimal assignment problem can be formulated and solved (using for example Hungarian algorithm) to find the best allocation matrix $X = \left[x_{ij}\right]_{|S|\times|S'|}$ of nodes from graphs *S* and S':

$$P(S,S') = \sum_{i=1}^{|W_G|}\sum_{j=1}^{|W_{G'}|} s_{ij}\cdot x_{ij} \to max \tag{8}$$

with constraints:

$$\sum_{i=1}^{|W_G|} x_{ij} \leq 1, j = \overline{1,|W_{G'}|} \tag{9}$$

$$\sum_{j=1}^{|W_{G'}|} x_{ij} \leq 1, i = \overline{1,|W_G|} \quad (10)$$

$$\bigwedge_{i\in\{1,\dots,|W_G|\}} \bigwedge_{j\in\{1,\dots,|W_{G'}|\}} x_{ij} \in \{0,1\}$$

The $P(S,S')$ is the value of structural similarity measure of graphs S and S'.

It is worth notice that analysed nodes could be part of the same or different graph. In the proposed method only shared nearest neighbours (first level neighbours) are taken into consideration. It can be easily extended to include neighbours of nearest neighbours (second level neighbours) and further. As another extension we could propose to include node neighbours weights (attributes) when similarity of nodes is calculated.

5. The concept of profile cloning detection

Let's introduce a following notation:
S – pattern network;
S' – analysed network;
$t_s \in [0,1]$ – threshold parameter for structural similarity of networks;
$t_{ID} \in [0,1]$ – threshold parameter for nodes global similarity;
w_s – person in analysed network that we are going to check if his identity was stolen nor not;
$P(S,S')$ – structural similarity of S and S';
i^{max} – maximum amount of iterations that algorithm should take;
ε – minimal allowed increase of structural similarity, that continues iteration of algorithm;
$t_s^{old} \in [0,1]$ – structural similarity value between S and S' from previous iteration.

We provide an approach to detect profile cloning using the method illustrated on Figure 5. The general procedure is as follows. After providing an input data, we are using a method to reveal hidden values of attributes for every node in analysed network. A mechanism how to do it was described in chapter 3. Next, we calculate a global node similarity function to check whether analysing network is similar enough to pattern network. If this similarity is low, then we can assume that in analysed network there are not enough information to detect a profile cloning or the structure of this network, in sense of social profile, is too different so that the possibility of cloning a profile is low.

In situation when the G_{ID} value is greater than defined threshold, we can consider another steps of method. In that case, we calculate a structural similarity measure between pattern network and structural network. If this value is high enough, then we can assume, that the personal profile of person w_s is nearly the same in analysed network than in pattern network. This situation could happen in two scenarios. First is when the considered person create an account in analysed OSN and started to make a relations with his friends. Another case is when someone creates an account in analysed network and tries to recreate a social profile of person w_s. Second situation appears when a profile cloning has place, so we can say that we detect profile cloning.

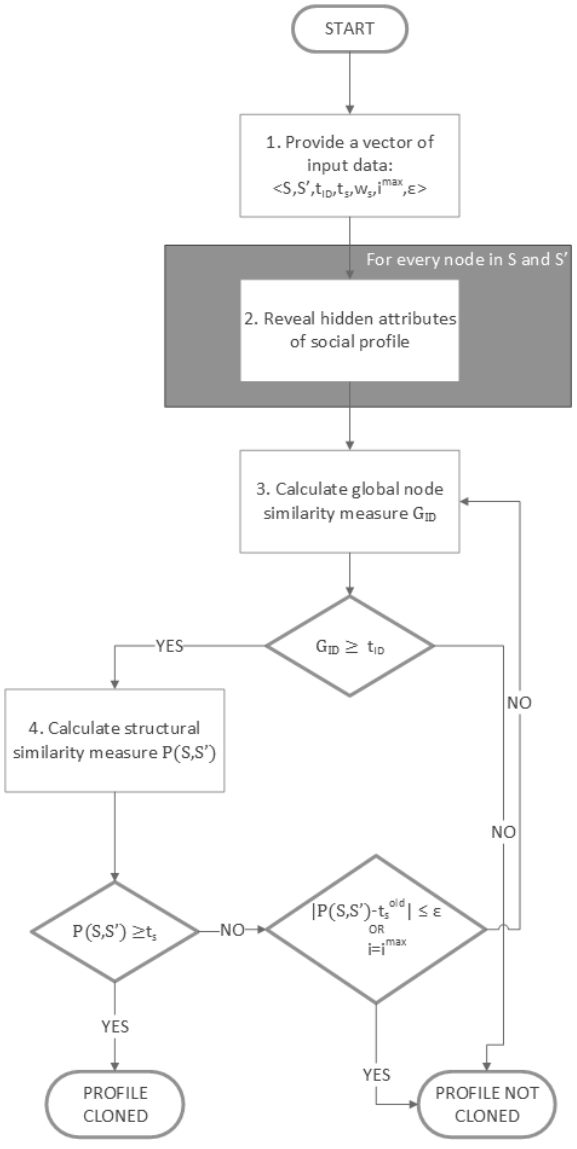

Fig. 5. Profile cloning detection algorithm

Other part of algorithm consists of conditions to stop the method when a profile cloning didn't have a place. For this purpose we use two approaches. First is to set an maximum for amount of iteration. This gives us an

opportunity to stop executing of method, but will also cause a risk, that algorithm will stop before detect a profile cloning. Another option used in our solution is to check whether an increasing of similarity measures are significant enough to continue searching. In that case, we assume that if we cannot obtain appropriate similarity of networks, then there is not any possibility that user's personal profile, in sense of attributes and relations between him and other people in OSN, was cloned. Because there is a risk that a second approach will not converge enough to stop executing an algorithm, a composition of first and second solution is used.

6. An Example

To show basic idea of presented approach, we used a simple social network *ExNet*, presented on Figure 6 with additional data, that defines persons social profile, described in Table 1.

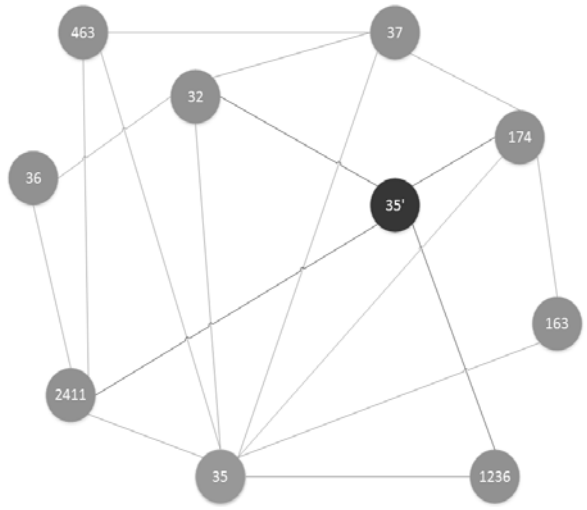

Fig. 6. Model of example network *ExNet*

In analysed case, person with ID=35 is a victim, while node with ID=35' represents hacker, who wants to steal 35's identity. Example shows, that part of the victim's social relationships are reconstructed.

Tab. 1. Definition of people's social profile in example network. Node with ID=35 is a victim and node with ID=35' is hacker

ID	Name	Gender	School	Degree	Employer	Position	Birth-day	Loca-tion	Relation-ships
32	Niko Parda	Female	Harvard University	PhD	East Man	Manager	1979	USA	Single
35	Sara Abraham	Female	Arcadia University	Master's	Owens	Web Developer	1980	USA	Single
35'	Sara Abraham	Female	Arcadia University	Master's	Owens	Web Developer	1980	USA	Single
36	Sara Abraha	Female	Carolina University	Master's	Owens	Web Developer	1980	USA	Single
174	David Ernox	Male	Michigan University	Master's	Qpass	Java Developer	1984	USA	Single
463	Sara Abram	Female	Michigan University	Master's	AppNet	Web Developer	1985	USA	Single
1236	Tom Banho	Male	Arcadia University	Bachelor	Xing	Network Manager	1979	USA	Married
2411	Rose Milan	Female	Koln University	PhD	Axvert	Manager	1972	USA	Single
163	Charls Selvin	Male	Pietersburg University	Bachelor	Sony	Accountant	1979	UK	Married
37	Silvia Jacson	Female	Carolina University	Bachelor	MySpace	Computer Data Clerk	1978	Australia	Married

Because all value of social profile attributes are known, we do not have to use a revealing hidden values of social profile attribute algorithm, which has been described in section 2. First of all, we will compute nodes attribute similarity to check whether they exceed the threshold. If it has place, then the node with node similarity value higher than threshold can be a possible cloned profile. For our calculations we adopted some assumptions:

- Because we are using an equation (3) to calculate node similarity we assume, that an importance factor is the same for all the nodes;
- We set the threshold value to 8.

Results of our work show Table 2.

Tab. 2. Results of nodes attribute similarity computation

Node	**32**	**35**	**35'**	**36**	**174**	**463**	**1236**	**2411**	**163**	**37**
32	10	3	3	3	2	3	3	5	1	1
35	3	10	**9**	7	3	5	2	3	0	1
35'	3	**9**	10	7	3	5	2	3	0	1
36	3	7	7	10	3	5	1	3	0	2
174	2	3	3	3	10	4	2	2	1	2
463	3	5	5	5	4	10	1	3	0	1
1236	3	2	2	1	2	1	10	1	4	2
2411	5	3	3	3	2	3	1	10	0	1
163	1	0	0	0	1	0	4	0	10	2
37	1	1	1	2	2	1	2	1	2	10

For example, calculating $ID(35,35')$ gives us:

$$(35,35')=p_{ID}(35,35')+p_{name}(35,35')+...=$$
$$=0+1+1+1+1+1+1+1+1+1=9$$

Because $(35,35')$ has got higher value than defined threshold, the node with ID=35' will be detected as a cloned profile. Figure 6 shows that this is exactly a hacker who tries to steal the identity of 35. In that case presented solution detects attacker without even using a structural similarity, which is another part of the algorithm and in connection with node attributes similarity gives us better precision of the whole approach.

For calculation of structural similarity of example network *ExNet*, we used equations described in section 4 of article. Using them, we obtained a results showed in Table 3 (matrix $P_{node}(S,S')$, where $S=S'=ExNet$).

Tab. 3. Results of graphs structural similarity computation

Node	**32**	**35**	**35'**	**36**	**174**	**463**	**1236**	**2411**	**163**	**37**
32	0,109	0,168	0,097	0,052	0,121	0,070	0,113	0,097	0,066	0,103
35	0,168	0,260	**0,149**	0,081	0,187	0,108	0,175	0,150	0,102	0,159
35'	0,097	**0,149**	0,086	0,047	0,107	0,062	0,101	0,086	0,058	0,091
36	0,052	0,081	0,047	0,025	0,058	0,034	0,054	0,047	0,032	0,050
174	0,121	0,187	0,107	0,058	0,134	0,077	0,126	0,108	0,073	0,114
463	0,070	0,108	0,062	0,034	0,077	0,045	0,072	0,062	0,042	0,066
1236	0,113	0,175	0,101	0,054	0,126	0,072	0,117	0,101	0,068	0,107
2411	0,097	0,150	0,086	0,047	0,108	0,062	0,101	0,087	0,059	0,092
163	0,066	0,102	0,058	0,032	0,073	0,042	0,068	0,059	0,040	0,062
37	0,103	0,159	0,091	0,050	0,114	0,066	0,107	0,092	0,062	0,097

In that case, the nodes with highest similarity value will be treated as the same person, which in our situation means that we detect a hacker, who will try to steal victim's identity. It is easy to notice that structural similarity for node 35' has the greatest value (equals 0,149) for node 35, so node 35' may be node 35' in sense of network's structure similarity. Again, in that way we were able to do a successful search for attacker in Online Social Network.

7. Conclusions

Presented approach shows that detecting a profile cloning using analytical methods is possible. This allows us to develop solutions, which gives us opportunity to automate the process of discovering identity stealing in Online Social Networks. Moreover, a revealing hidden attributes values model, which is a part of detecting profile cloning procedure, can be also useful in analysing networks with lack of information about users and gives us also opportunity to analyse a procedure used by attackers to detect hidden attribute values. If we know how hackers are working to steal people identity, we can define vulnerabilities of their action, which helps us to design methods to prevent from an unauthorized access to our attribute values.

Despite the effectiveness of considered approach, it is still place to extend presented method. For example, it will be good to design specialized structural similarity measures dedicated directly for Online Social Networks, which take into account a social profile characteristics and attributes. In our future work, we also plan to add ability to predict creation of relation between people in considered OSN. An example of predictor for making relations between people in Online Social Networks is described in details in [16]. This allows us to forecast an incident of profile cloning before this process will be completed by attacker. Thanks that, we will be able to prevent hacker from stealing user identity, which will definitely increase a value of implemented solution.

Another problem which should be taken into future consideration is performance of approach for big networks [15]. Currently, we have to calculate node attribute similarity for all pair of nodes, which is computationally complex. To prevent this, we can detect some patterns or communities and focus only on them, because another set of nodes will be not similar enough. However at present time this is only a hypothesis and should be designed and tested, especially on big networks.

It will be good also to develop more complex similarity measures between attribute values, because binary approach presented in article should be not precisely enough for continuous values. In that case, we could use for example an Euclidean metric. So our next step in future work will be adapting some well-known metrics to our attribute values similarity model. It is also possible to take into account node's attribute similarity and node's structural similarity (as two criteria) in multicriteria approach for profile cloning detection.

8. Bibliography

[1] Faith Cranor L., "Internet privacy", *Communications of the ACM*, Vol. 42, 28–38 (1999).

[2] Mo M., Wang D., Li B., Hong D., King I., "Exploit of Online Social Networks with Semi-Supervised Learning", "The 2010 International Joint Conference on Neural Networks (IJCNN)", 18–23 July 2010, Barcelona, Spain.

[3] Khayyambashi M., Rizi F., "An approach for detecting profile cloning in online social networks", 7th International Conference on "e-Commerce in Developing Countries: With Focus on e-Security (ECDC)", 17–18 April 2013, Kish Island, Iran.

[4] Jeff Smith H., *Managing Privacy*: *Information Technology and Corporate America*, UNC Press Books, University of North Carolina Press, 1994.

[5] Kontaxis G., Polakis I., Ioannidis S., Markatos E.P., "Detecting social network profile cloning", IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 21–25 March 2011, Seattle, WA, USA.

[6] Zhu X., Ghahramani Z., Lafferty J., "Semi--Supervised Learning Using Gaussian Fields and Harmonic Functions",

in: *Proceedings of the Twentieth International Conference on Machine Learning*, 912–919, Washington, DC, 2003.
[7] Blum A., Chawla Sh., *Learning from Labeled and Unlabeled Data using Graph Mincuts*, Computer Science Department, paper 163, Carnegie Mellon University, Pittsburgh, USA, 2001.
[8] Bilge L., Strufe T., Balzarotti D., Kirda E., "All your contacts are belong to us: Automated identity theft attacks on social networks", in: *Proceedings of the 18th International Conference on World Wide Web* (*WWW 2009*), 551–560, 20–24 April, 2009, Madrit, Spain.
[9] Tarapata Z., Kasprzyk R., "An application of multicriteria weighted graph similarity method to social networks analyzing", in: *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*, Athens (Greece), 366–368, IEEE Computer Society, 2009.
[10] Blondel V., Gajardo A., Heymans M., Senellart P., Van Dooren P., "A Measure Of Similarity Between Graph Vertices: Applications To Synonym Extraction And Web Searching", *SIAM Review*, Vol. 46, No. 4, 647–666 (2004).
[11] Nikolić M., "Measuring similarity of graph nodes by neighbor matching", *Intelligent Data Analysis*, Vol. 16, No. 6, 865–878, (2012).
[12] Jeh G., Widom J., "SimRank: a measure of structural-context similarity", in: *Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 538–543, Edmonton, 2002.
[13] Tarapata Z., "Multicriteria weighted graphs similarity and its application for decision situation pattern matching problem", in: *Proceedings of the 13th IEEE/IFAC International Conference on Methods and Models in Automation and Robotics*, 1149−1155, Szczecin, Poland 2007.
[14] Zager L., *Graph similarity and matching*, PhD thesis, MIT, 2005.
[15] Bartosiak C., Kasprzyk R., Tarapata Z., "Application of Graphs and Networks Similarity Measures for Analyzing Complex Networks", *Biuletyn Instytutu Systemów Informatycznych*, Vol. 7, 1–7, (2011).
[16] Liben-Nowell D., Kleinberg J., "The Link--Prediction Problem for Social Networks", *Journal of the American Society for Information Science and Technology*, Vol. 58, No. 7, 1019–1031 (2007).
[17] Zabielski M., Kasprzyk R., Tarapata Z., Szkółka K., "Methods of Profile Cloning Detection in Online Social Networks", in: *Proceedings of the 20th International Conference on Circuits, Systems, Communications and Computers* (*CSCC 2016*), Corfu Island (Greece), July 14–17, 2016.
[18] Fire M., Goldschmidt R., Elovici Y., "Online Social Networks: Threats and Solutions", in: *IEEE Communications Surveys & Tutorials*, Vol. 16, No. 4, 2019–2036 (2014).
[19] www.statista.com

Wykrywanie klonowania profili w internetowych sieciach społecznych

M. ZABIELSKI, Z. TARAPATA, R. KASPRZYK, K. SZKÓŁKA

Zagadnienie ochrony prywatności w Internecie istnieje od dosyć dawna, jednakże wraz z pojawieniem się internetowych sieci społecznych znaczenie tego tematu wzrosło drastycznie. Wynika to z faktu, iż sieci te są źródłem istotnych informacji osobowych powiązanych z konkretnym człowiekiem, które w dość prosty sposób są możliwe do wprowadzenia do informacji publicznej. Ważną kwestią jest zatem opracowanie mechanizmu, który uniemożliwi osobom niepowołanym wykrycie danych osobowych ukrytych przed dostępem publicznym. W pracy podjęto próbę przedstawienia koncepcji mechanizmu wykrywania klonowania profilu użytkownika w Internetowej Sieci Społecznej z wykorzystaniem teorii grafów i sieci. Analizując podobieństwo strukturalne sieci wraz z atrybutami opisującymi osobę w niej, jesteśmy w stanie znaleźć osoby próbujące ukraść naszą tożsamość.

Słowa kluczowe: wykrywanie klonowania profilu, internetowe sieci społeczne, naruszenie prywatności danych.