Виявлення Фейкових Облікових Записів у Соціальній Мережі «Facebook»

Олеся Войтович, Андрій Дудатьєв
кафедра захисту інформації
Вінницький національний технічний університет
Вінниця, Україна
voytovych.olesya@vntu.edu.ua, dudatyev.av@gmail.com

Віталій Головенько
кафедра захисту інформації
Вінницький національний технічний університет
Вінниця, Україна
torvald124@gmail.com

Fake Accounts Detection in Social Network «Facebook»

Olesia Voitovych, Andrii Dudatyev
dept. of Cybersecurity
Vinnytsia National Technical University
Vinnytsia, Ukraine
voytovych.olesya@vntu.edu.ua, dudatyev.av@gmail.com

Vitalii Holovenko
dept. of Cybersecurity
Vinnytsia National Technical University
Vinnytsia, Ukraine
torvald124@gmail.com

***Анотація*—У статті запропоновано ознаки фейкових облікових записів у соціальній мережі «Facebook». На основі запропонованих ознак та використанні рейтингових оцінок розроблено систему підтримку прийняття рішень при виявленні фейкових облікових записів.**

***Abstract*—The fake accounts` attributes in social network «Facebook» are proposed in the article. The proposed attributes as well as rating scores are used for decision support system for fake accounts detection developing.**

Ключові слова—фейкові облікові записи; соціальні мережі; кібербезпека; інформаційна війна; рейтингові оцінки

Keywords—fake accounts; social networks; cyber security; information warfare; rating scores

I. INTRODUCTION

Social networks are specific places for doing special information operations especially informational psychological operations targeted on society [1, 2]. Hundred millions of people around the world use social networks for communication, reading news and so on. However, great amount of people use social networks as a tool for manipulation of individual and sociable mind by using informational throws-in (mems) [3]. For manipulation people use fake accounts in which there is no information about them or there is false information on their profile. Using of fake accounts is usually targeted at changing sociable mind in one form or another and it doesn`t matter the aims of people who create fake accounts. [4].

II. DISTINGUISHING OF FAKE ACCOUNTS ATTRIBUTES

The research [5-11] showed that the basic categories of fake account attributes such as likes, personal information, statuses and posts, friends, photos can be distinguished (fig. 1).

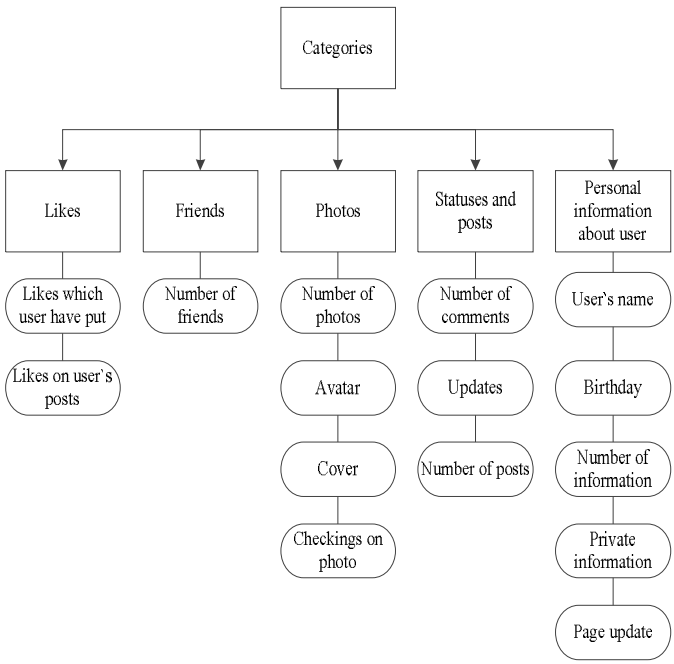

Fig. 1. Scheme of categories of fake account attributes

Likes [8] can be distinguished by the attributes of their quantity and people who leaved likes on a page.Likes can be leaved by friends or strangers. Quantity of likes also has meaning for indicating fakes. If a user has more likes on a post than quantity his/her friends, it could mean that user has got that likes illegally. Lack of likes on a page also shows the user`s isolation and as a result – fake.

Parameters of the fake attributes model in a category "Likes" are written in a form of tuples:

LIKES = {FROM; QUANTITY}

Personal information on a page [8-9] can show a lot about fake or real status of user. Personal information can be shared on user`s name, birthday, quantity of information about user, contradictory information and private information for future analysis.

Birthday has attributes that show fakeness of account. Fake users often don`t mind about detailed filling of their pages and leave birthday as default (usually the 1st of January). There is a possible situation when user`s age is doubtful or doesn`t match with other dates on the page. For example, user is fifteen but other information on the page says that user leaved a universityten years ago.

User`s name is difficult to analyze because a lot of people with the same name and surname exist. But user`s name should be checked if it matches with a names of celebrities. Also it should be checked if user`s name is typical for user`s country.

Lack or minor amount of personal information on user`s profile proves that a user doesn`t want him/her to be identified by other users. So it`s a prove of fake account.

Contradictory information on a page is one of the authentic attributes of fakes but it needs difficult analyzing. For example, post information does not correspond to profile information or user`s groups do not correspond to user`s interests.

E-mail and mobile phone number also concerns to personal information about user. Users seldom put such private information in open access except of fake accounts or special advertising accounts.

Parameters of fake attributes model in category "Personal information about user" are written in a form of tuples:

PERSONAL INFORMATION ABOUT USER = {DATE OF BIRTH; USER NAME; NUMBER OF INFORMATION; CONTRADICTORY INFORMATION; PRIVATE INFORMATION}

Statuses and posts on a page should be analyzed as one because their difference is only in their location on the page. They can be analyzed by such attributes: update/create frequency and comments. Statuses and posts are usually used for advertisement [10].

Posts and statuses update/create frequency indicates a user`s activity. If posts or statuses are created seldom or too often, it`s one of the fake sings. If a user added a post/status a long time ago and doesn`t update it during long period of time, there is a probability that the account is fake.

Quantity of comments also shows a profile`s activity. Lack or great amount of comments is usually belonged to fake accounts. Comments can be leaved by user`s friends or strangers.

Parameters of fake attributes model in a category "Statuses and posts" are written in a form of tuples:

STATUSES AND POSTS ON PAGE = {ADVERTISING; UPDATES; COMMENTS}

User`sfriends` analyzingisveryimportantforfakedetectionbecauseitshows the profileactivityinasocialnetworkanduser`s interests [11].

It`s difficult to analyze fakenessdependence of user`s friendsbecause it`s necessary to analyze friends themselves to make right conclusion about fakeness of a profile. For example, if a user has friends that are fakes, there is a possibility that the user is a fake. If a user doesn`t have friends, there is a great possibility that his/her profile is used for purposes other than communication with people. Big amount of friends got for short period of time after profile creation cause suspicions, so that profile is probably fake.

Parameters of fake attributes model in a category "Friends" are written in a form of tuples:

FRIENDS = {NUMBER OF FRIENDS; INFORMATION ABOUT FRIENDS}

User`s photos` analyzing is also very important and at the same time the most difficult part of the fake accountsanalysis. Firstly, lack of photos on avatar and in albums means that this profile is a fake. Secondly, if there are photos on a page, they should be analyzed anyway. Those photos can match with other pictures in the Internet or with other users` photos. A user can upload photos of celebrities, animals, other objects instead of his/her real photos. Quantity of photos is also an important attribute because if there is large or small number of photos, it means that the account is a fake, or that the user is not active appropriately.

Parameters of fake attributes model in a category "Photos" are written in a form of tuples:

PHOTO = {PHOTOES ON PROFILE; AVATAR; COVER}

Of course, separately these criteria cannot point on the fakeness of a profile clearly because system of criteria analysis only can questionthecertainty of account.

For a more trustworthy definition of profile status it`s necessary to use the analysis with as large as possible number of criteria [12-13].

In this article other important parameters of accounts aren`t considered such as page creation time, speed of friends addingand connections with each other. These and other parameters will be considered in a future research.

III. RATINGDECISIONSUPPORTSYSTEM

Fordecisionsystemaboutfakeness ofaccounttheratingscoresmethodwasproposedthatallowstaking into account the weight coefficients of parameters` significance andevaluateinformationthatisdividedintocategories [7].

Let the system evaluates by *n*parameters, x_i – values of*i*parameter. The representation of the rating system is a linear convolution, the mathematical model of which is written in the formbelow (1) [14]:

$$F = \sum_{i=1}^{n} \lambda_i x_i , \quad (1)$$

where λ_i is a weight of x_i parameter, which is determined by an expert.

Based on the rating approach and multivariate analysis, a group of rating assessments of indicators is developed and a link between them is established. The application of the rating approach implies that rankings are assigned to all groups of factors.

Let the system of evaluated attributes be described on the basis of a given set of indicators, such as $X = (x_1, \ldots, x_i, \ldots, x_n)$. Indicators may be heterogeneous: numeric, logical, lexical, vector, etc. To operate heterogeneous indicators for each of them a normalized function is introduced, which any value of x_i values translates into a set of real values on the segment [0; 1], then $0 \leq x_i \leq 1$. Rationing can lead to overestimation or underestimation of the actual indicator, but this negative effect is neutralized by inputting a weighting factor for each indicator, which is determined by the empirical method (expert estimation method). If some dependence of the indicators among themselves exists, it is necessary to take into account the mixed constructions, where coefficients are the coefficients of correlation of the corresponding pair.

If weighing ratios are selected under normalization, then the target function will act as a ranking on the appropriate level of the system hierarchy. In order that the process of rating approach has the maximum effect, the indicators of all factors must be involved. This condition is automatically executed when the rating score for each level of the hierarchy coincides with the target function. Integral ranking reflects the priorities of the indicators. The formation of these indicators, and hence the formation of a ranking, in this paper is carried out by experts.

IV. EXPERIMENTAL STUDIES

To work with the social networking data in Facebook, the Python programming language and the Facebook-SDK library were selected [15]. In order to gain access to user information on the Facebook social network, one must obtain an authentication token that specifies the rights of the developer to access the data in Facebook. Fig. 2 shows the process of getting data about user from social network «Facebook» using Facebook-SDK library.

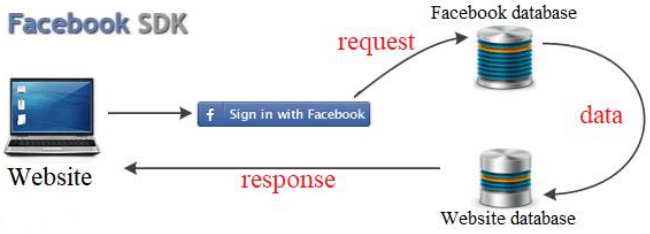

Fig. 2. Algorithm of getting data for research from social network «Facebook»

The designed software consists of modules that read the user information and analyze it. Modules were developed to allow reading and processing information about user's friends, profile photos, photo tags, background photos, number of posts, user's birthday, personal information, page updates, user name, likes on posts and likes, that the user has placed (fig. 1).

As an indicator, a system of scores is selected, indicating that the user account is a fake one [6-7]. Each of the parameters in the analysis receives a certain number of points from 1 to 5. So, 100 points shows that the account is a fake, and 0 points shows that it is a real one. Different weight coefficients from 1 to 3 were chosen for different categories, based on expert knowledge. If the result of the research is from 10 to 45 points, then the system decides that the account is real, from 55 to 100 points - the account is a fake. However, if the result is in the range of 45 to 55 points - further studies should be conducted. As a result, the output and detailed information about the criteria that influenced the outcome are displayed on the screen. Fig.3 shows the application window indicating that the account is a fake. Fakeness is indicated by the lack of information about user, the lack of likes and photo tags, as well as the absence of likes on the very few user posts.

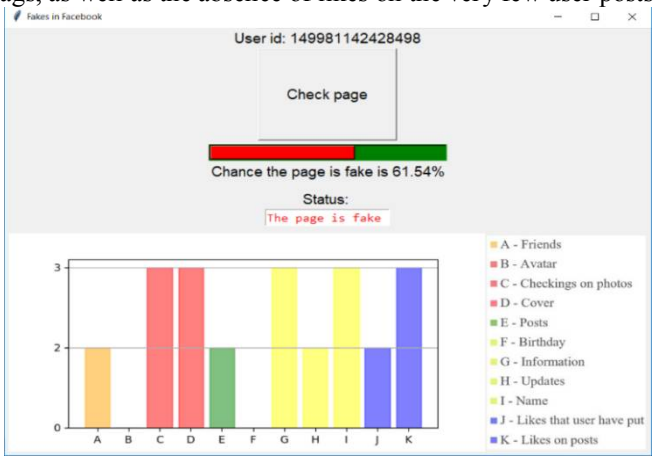

Fig. 3. Look of program window with detected fake account

Fig. 4 shows the application window indicating that the user turned out to be real. Indicators of fakeness in this case are incompletely filled account profile and a little number of likes, which is quite normal for real users, that is why the significant criteria of fakeness did not affect the result.

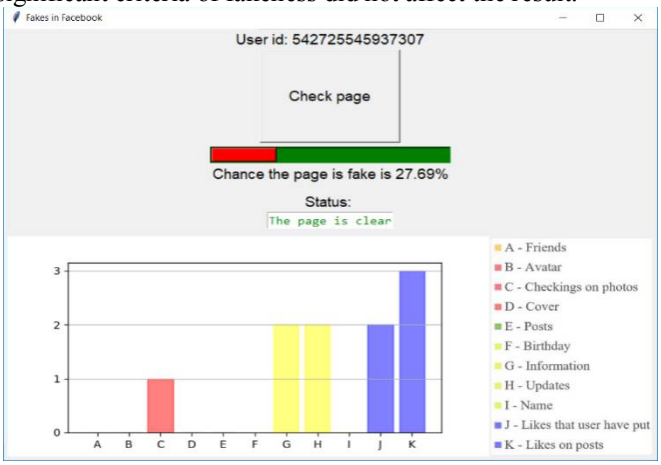

Fig. 4. Look of program window with detected real account

For objective testing of the software, various user accounts were checked on the Facebook social network, which included both fake and real accounts. The results of analysis of 21 accounts are shown in the table 1.

TABLE I. THE RESULTS OF ANALYSIS OF 21 ACCOUNTS

User	*Account status*	*Program result, points*	*Program conclusion*
Vitalii Holovenko	Real	37,8	Real
Татьяна Головенько	Real	45,9	Not defined
OleksandrTopchii	Fake	67,5	Fake
IvanVorobyov	Real	19,8	Real
Alex Rudyk	Fake	90,0	Fake
Ольга Гнатюк	Real	21,6	Real
Петро Петрович	Fake	96,3	Fake
Andrii Beatle	Fake	62,1	Fake
Жека Олейник	Real	34,2	Real
Владислав Круговой	Real	54,0	Not defined
Олеся Войтович	Real	48,6	Not defined
Talii Santie	Fake	77,4	Fake
Jenny Rahl	Fake	70.2	Fake
Sergey Hubchakevych	Real	28,8	Real
Георгий Выфв	Fake	87,3	Fake
Alice Black	Fake	66,6	Fake
Liliana Vess	Fake	69,3	Fake
Konrad Von H.	Fake	67,5	Fake
Сергей Тарахта	Real	24,3	Real
Иван Петров	Fake	44,8	Real
Fin Age	Fake	65,7	Fake

Taking into account the results obtained, we can conclude the reliability of the results obtained by the software is 81%.

CONCLUSION

Main attributes of accounts in social network «Facebook» that allows to detect fake accounts were considered and analyzed. Each of attributes by their possible parameters and influence on account status were analyzed. Each of attributes is divided into categories and appropriate tuples were formed.

Attribute model for detection fake accounts that includes categories such as likes, personal information, statuses and posts, friends, photos was proposed.

Rating decision support system for fake accounts detection in social network «Facebook»is developed. Experimental research showsthe decision support system certainty is 0,8.

To improve certainty of decision support system it`s planned to analyze more social network`s parameters and reduce the dimensionality of the feature space.

REFERENCES

[1] Voitovych O., Holovenko V. Research of social networks as a source of information in warfare. Inżynier XXI wieku projectujemy przyszlosc: monografia / pod red: Jacek Rysiński. Bielsko-Biała, 2016. C. 111-119.

[2] Дудатьєв А. В., Войтович О. П. Інформаційна безпека соціотехнічних систем: Модель інформаційного впливу. Інформаційні технології та комп'ютерна інженерія. 2017. С.16-21.

[3] Коршунов А., Белобородов И., Бузун Н. Анализ социальных сетей: методы и приложения. Труды Института системного программирования РАН. 2014. Т. 26. № 1. С. 439-456.

[4] Дудатьєв А. В. Комплексна інформаційна безпека СТС: моделі впливу та захисту : монографія. Вінниця: ВНТУ, 2017. 128 с.

[5] Нежданов И. Ю. Технологии информационных войн в Интернете URL: http://bash.rosmu.ru/activity/attach/events/1283/01.pdf (дата звернення: 10.01.2017)

[6] Войтович О. П., Буда А. Г., Головенько В. О. Дослідженння методів аналізу соціальних мереж як середовища інформаційних війн / Войтович О. П. //Тези доповідей Шостої Міжнародної науково-практичної конференції «Методи та засоби кодування, захисту й ущільнення інформації» м. Вінниця, 24-25 жовтня 2017 року. – Вінниця: ВНТУ, С. 67-70 - 2017.

[7] Войтович О. П., Дудатьєв А. В., Головенько В. О. Модель та засіб для виявлення фейкових облікових записів у соціальних мережах // Вчені записки Таврійського національного університету ім. В. І. Вернадського. Серія: Технічні науки. Частина 1 - 2018. - № 1 Том 29 (68). – с. 112 – 119.

[8] Губанов Д. А., Новиков Д. А., Чхартишвили А. Г. Социальные сети: модели информационного влияния, управления и противоборства. М.: Физматлит, 2010. 228 с.

[9] Michal Kosinski, Sandra C. Matz, Samuel D. Gosling, Vesselin Popov, David Stillwe. Facebook as a Research Tool for the Social Sciences. Opportunities, Challenges, Ethical Considerations, and Practical Guidelines. American Psychologist. 2015. Vol. 70. No. 6. 543-556 pp.

[10] Aaron Aguis. 10 Metrics to Track for Social Media Success. URL: https://www.socialmediaexaminer.com/10-metrics-to-track-for-social-media-success/ (дата звернення: 10.01.2017)

[11] Батура Т. В., Копылова Н. С., Мурзин Ф. А., Проскуряков А. В. Методы анализа данных из социальных сетей. Вестник НГУ. Серия: Информационные технологии. 2013. Т. 11. Вып. 3. С. 5-21.

[12] Берни Хоган. Анализ социальных сетей в интернете, 2013. URL: https://postnauka.ru/longreads/20259 (дата звернення: 10.01.2017)

[13] Горчинская О., Ривкин А. Анализ данных социальных сетей. Открытые системы. СУБД. 2015. №3. С. 22-23.

[14] Худяков Ю. Г., Николайкин Н. И., Андрусов В. Э. Управление опасностями производственной среды: Монография. М.: ООО «Проспект», 2017. 122 с.

[15] API Reference URL: https://facebook-sdk.readthedocs.io/en/latest/api.html (дата звернення: 10.01.2018)