TELKOMNIKA, Vol.13, No.2, June 2015, pp. 373~380
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v13i2.1436

An Introduction to Journal Phishings and Their Detection Approach

Mehdi Dadkhah*[1], Tole Sutikno[2], Mohammad Davarpanah Jazi[3], Deris Stiawan[4]
[1]Department of Computer and Information Technology, Foulad Institute of Technology
Foulad shahr, Isfahan 8491663763, Iran
[2]Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[3]Department of Computer and Information Technology, Foulad Institute of Technology
Foulad shahr, Isfahan 8491663763, Iran
[4]Department of Computer System Engineering, Universitas Sriwijaya, Palembang, Indonesia
*Corresponding author; e-mail: dadkhah80@gmail.com[1], tole@ee.uad.ac.id[2]

Abstract

Nowadays, the most important risk and challenge in online system are online scam and phishing attacks. Phishing attacks have been always used to steal important information of users. In this kind of scam, attacker direct victim to fake pages using social engineering techniques, then, starts stealing users` important information such as passwords. In order to confronting these attacks, numerous techniques have been invented which have the ability to confront different kinds of these attacks. Our goal in this paper is to introducing new kind of phishing attacks which are not identifiable by techniques and methods which have been invented to confronting phishing attacks. Unlike other kinds of phishing attacks which target all kinds of users, researchers are the victims of these kinds of journal phishing attacks. Finally, we`ll introduce an approach based on classification algorithms to identify these kind of journal phishing attacks and then we`ll check our suggested approach in error rate.

Keywords*: Phishing, Hijacked journal, Classification, Data Mining*

1. Introduction

Phishing attacks is an effort for accessing people important information like; username, password, and credit cards information, using social engineering techniques [1]. These attacks were explained in 1987 by details and were used in 1996 for the first time [2]. In these attacks, to increasing success ratio, attackers try to represent themselves in a way that victims trust them and accept them as a legal agents of valid organiztion such as banks. In these kinds of attacks, phishers (forgers who use phishing attacks), begins their plan by designing a website which is similar to legal website. Having done this step, the must find a way to persuade their victims to enter their own website and enter his/her information. So, main target on a phishing attack is to use a fake connection which begins with a e-mail including fake URL from a banks or governmental agency. Attacker or phishing attack designer tries to use cases which are attractive to victims and can pay their attention. Then tries to achieve name, phone number or any other kinds of information which can be used for advancing his goals, on the other hand phishing attacks are used to steal victims` identification using computer networks. These kinds of attacks are designed generally by means of accessing to people IDs and passwords. But generally, includes any kind of information which illegal use of them will follow attackers benefits.

Many studies and efforts have been done to introducing different kinds of phishing attacks and their confronting ways. Generally, different kinds of phishing attacks include deceptive phishing [3], phishing based on destructive software [4], web trojans [5], pharming [6], phishing injection [7], phishing using fake applications [8], domain hijacking [9], spear phishing [10] and changing user system settings attacks [6]. To confront these attacks many techniques and methods have been invented such Sign-in Seal [11], developing expert system based on characteristics of web pages in order to detect phishing websites [12], genetic algorithm based on anti-phishing techniques [13], detection of phishing attacks based on categorizing super links [14], attribute-based prevention of phishing attacks [15], content based on anti-phishing approach [16], confronting phishing attacks by two step identification [17], detection of phishing

Received January 6, 2015; Revised February 9, 2015; Accepted February 20, 2015

pages based on associated relationships [18], detection of phishing pages by comparing the amount of difference between the address string and the white list [19], ranking based anti-phishing approach [20], and using data mining algorithms [21]. Mentioned methods and techniques identify different kinds of phishing attacks by 27 recognized features but are useless against journal phishings which we want to introduce because 27 key features have been gotten from most related websites with e-commerce. In [22, 23] these attacks have been addressed as hijacked journals and some features of these kind of phishing attacks have been mentioned and general guides on this have been given to researchers. In [24, 25] discussions about fake publisher and open access publisher were been taken care for, but a definite confronting way has not been suggested and only few guides have been given to researchers about predatory publishers. Our goal on this paper is to extract related features by these kinds of phishing attacks and present a method to detecting and confronting with them.

2. Introduction to Journal Phishings

In this paper we`ll introduce journal phishings and try to extract these kinds of phishing attacks feature and finally present a approach based on classified algorithms to confront them. As mentioned before, in phishing attacks, phishers deceive victims by designing a fake website which is similar to the original and using social engineering techniques then steal victim sensitive information including password by directing them into fake websites and financial resources. Fake journals works with name and credit of some valid journals but in fact have no relationship with those journals and like phishing attacks, follow financial motivations with this difference that in this kind of scamming, the victim doesn`t give his/her sensitive information to forgers but delivers financial resources directly to them. In this kind of attack which we will be called "journal phishings" so on, the forgers mostly deceive their victim who are mostly researchers, by designing a fake web page and using valid journals name and ISSN. The forgers go after journals which are active in print version and by designing a website with original journal features start scamming from researchers and by receiving high sums, they will publish the victims papers. In these kinds of phishing attacks as deceptive phishing attacks, social engineering is exclusively used. The process of a journal phishings attack is shown in Figure 1.

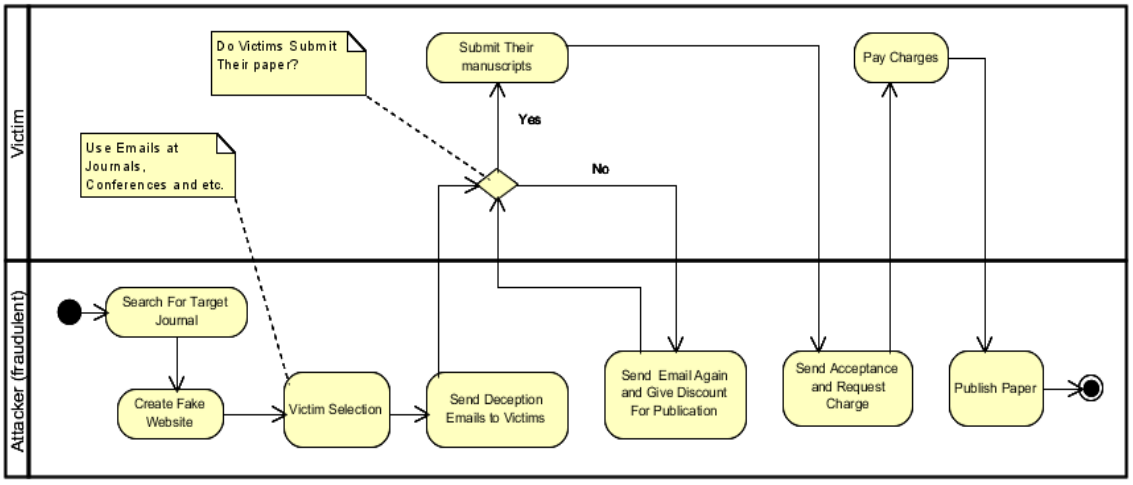


Figure1. The process of a journal phishing attack

Table 1 also shows common features between phishing attacks and hijacked journals and justifies naming hijacked journals as one of the phishing attacks.

Table 1. Common features between phishing attacks and journal phishings (Hijacked journals)

Feature	Journal phishing (high jacked journals)	Phishing attacks
Using social engineering	Very high	Average
Having financial motivations	Yes	Yes
Sending e-mail to deceive the victims	Yes	Yes
Using same name or domain	Yes	Yes
Choosing victim	Yes	In spear phishing attacks certain victims are attentional
Short life time of designed fake website	Fake journals websites are usually available for a short period of time	Phishing websites are usually available for a short period
Using available weaknesses in internet protocols like TCP/IP	Yes	Yes

3. Identifying Journal Phishing by Classification Algorithm

To classifying data and mechanism choices process classifying algorithm can be used. Classifying algorithm is applied to classifying data and extracting the sample from a set of data. By classifying algorithm, a sample can be extracted from a set of data, then use extracted sample for making decision about future data. Classifying algorithms have different types which we can name C5, CHAID, QUEST and C&R tree as the most popular. Extracted samples from these algorithms are mostly as a decision tree [26].

We need data to extract related journal phishings features, so we should collect a list of known journal phishings which have been detected. We study this list from academic resources which can be provided. According From our observation on collected journal phishing websites, key features of these phishing attacks will be extracted. Table 2 presents these features with measurable amount for each.

Table 2. Used features to recognizing phishing journals.

Rank	Adjective Name	Kind	Measures
1	Domain ranking	Logical	1=Having page rank 0=Not having page rank
2	Using external links	Discrete	L=Numbers of external link less than 2 M=Numbers of external links between 2 and 7 H=Numbers of external links more than 7
3	Domain lifetime	Logical	0=Short lifetime 1=Long lifetime
4	Indexing in popular databases	Logical	1=Indexed 0=Not indexed
5	Sequence in searching results	Discrete	L=Contained first 2 results M=Contained 2 to 4 results H=Other results
6	Entered countries to journal website	Discrete	H=Among 1 to 4 countries M=Among 4 to 8 countries L=More than 8 countries NA=No information
7	Availablity of previous issues	Logical	1=available 0=Not available
8	Long URL	Logical	1=Long URL 0= Suitable URL
9	Journal aim and scope	Logical	1=General aim and scope 0=Specific aim and scop

3.1. Domain ranking

This feature will be checked in a relation with website domain. Because journal phishings are the copy of the legal website, so they don`t have high ranks in search engines. But, this feature is not correct all the time, because, it might be a journal without website and search engines detect fake website instead of the legal one (like hijacked Jokull journal or

www.jokulljournal.com which has rank in Google search engine). In mentioned method, we use Google search engine because of the ability to ranking website based on page rank. This feature is concerned as a Boolean variable in a way that if checking website has ranking, the variable will be 1 otherwise it will be 0.

3.2. Using external links

This feature concentrates on checked websites codes structure. In the case that external links provide images with checked website content, the website is suspected to be journal phishing because most of journal phishings use other websites copied content.

3.3. Domain lifetime

According to our survey on journal phishings, most of these websites domain have been registered few months before designing the fake website while some papers according to many years ago are available in journal archive. So, by using Whios databases, we can extract the amount of this feature and get on detecting phishing journals. Suitable lifetime is measured according to the first issue in journals.

3.4. Indexing in popular databases

In general, indexed journals in a popular database are interested and have value to victims and can attract them. One of these indexing database is Thomson-Reuters. Almost all of the journal phishings are detected are indexing in this data base. But, we should consider correct database because Cite Factor (http://www.citefactor.org) has indexed almost all of the hijacked journals with fake addresses and it`s not suitable for surveying.

3.5. Sequence in searching results

This feature has been added to increase accuracy on detecting phishing pages in mentioned method. In this method the title of the concerned journal has been searching on search engine and the website address will be reversed. This feature will be used to detect journal phishings which their legal one has electronic version.

3.6. Entered countries to journal website

According to our studies on available journal phishing websites it has been detected that each journal phishing victim belongs to a certain country or includes limited population. Therefore journal phishings can be detected by Alexa database (http://www.alexa.com) and classifying the website guests based on the country.

3.7. Availablity of previous issues

Previous issues in journal phishings are not usually available or just some of them are. Phishers prevent user accessing previous issue by designing a login page for accessing previous issue or mentioning writers` names or the papers subjects. The reason of this is that designing a website with previous issue gets a lot of time and sometimes because the forger doesn't access all previous issues.

3.8. Long URL

Some journal phishings use a long URL. Long URLs are usually used to hide doubtful parts on address bar. According to scientific principles, there is no standard length to detecting legal URLs from illegal ones but normally, if a URL seems long this might belong to a phishing website.

3.9. Journal aim and scope

Most of the journal phishings are in a way that accept papers with different subjects or have general aim and scope. Most of these journals have specific names which don`t represent subject domain (like Walia or http://www.waliaj.com) or their subjects are in a way that conclude different research fields (like Journal of technology).

In order to choose the suitable algorithm to detect journal phishings, first we need to provide a training dataset of journal phishings and the amount of 9 mentioned features in table 2 for each phishing website need to be measured, then classified algorithms be applied on provided training dataset and according to error ratio, the most suitable algorithm will be

chosen. Used training dataset must include phishing websites in addition to legal and main ones to be able to detect original websites too. We use IBM SPSS Modeler data mining tools for classifying data. According to finished feature Selection analysis, the possibility of each feature in detecting phishing attacks has been represented in Table 3. All of the features with first priorities can be selected as the tree root. This is important, because it is possible that one of the feature in the site can`t be measured and decision can be made by using different feature as the tree root. It is important to say that if the root feature changes, the decision tree will change.

Table 3. The effect of the each feature on detecting journal phishings

Feature	Important
Domain lifetime	1
Availablity of previous issues	1
Domain ranking	1
Journal aim and scope	1
Entered countries to journal website	1
Indexing in popular databases	0.997
Sequence in searching results	0.964
Using external links	0.85
Long URL	0.838

Figure 2 represents different algorithms error ratio. So, according to this error ratio, C5 algorithm will be the most suitable algorithm for detecting journal phishing attacks.

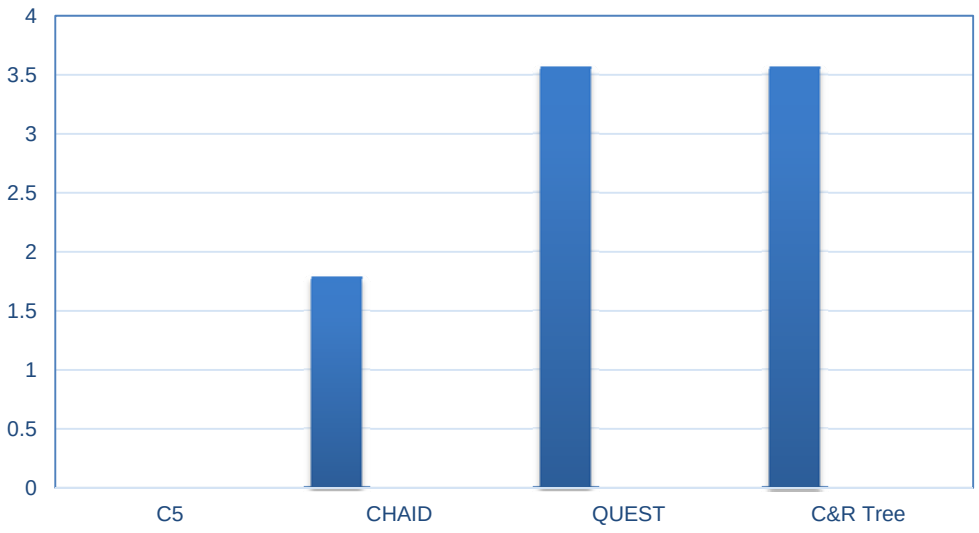
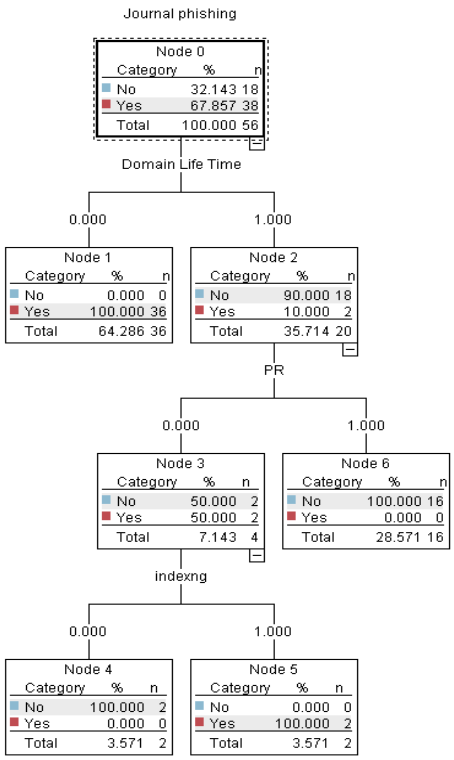
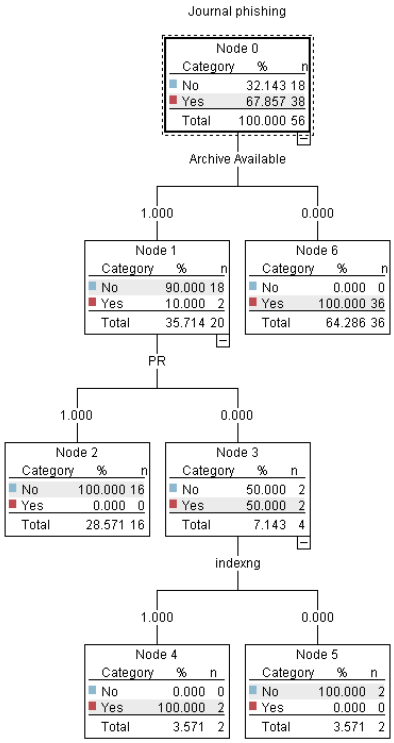


Figure 2. Algorithms with their error ratio in classifying data

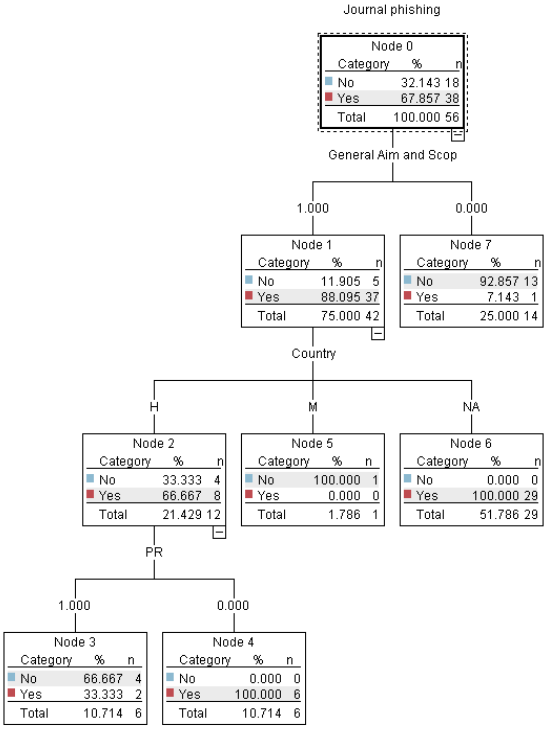
Meanwhile, Figure 3 shows decision trees in a case that different features are selected as root. But we should consider this if root feature changes, error ratio will change a little but the amount of these changes are low. We used IBM SPSS Modeler application to analyze data and extract decision tree based on our gathered data and built different decision trees by choosing different features as the root. This is important that the value of root feature may not be measurable and choosing another feature can do making decision as the root.



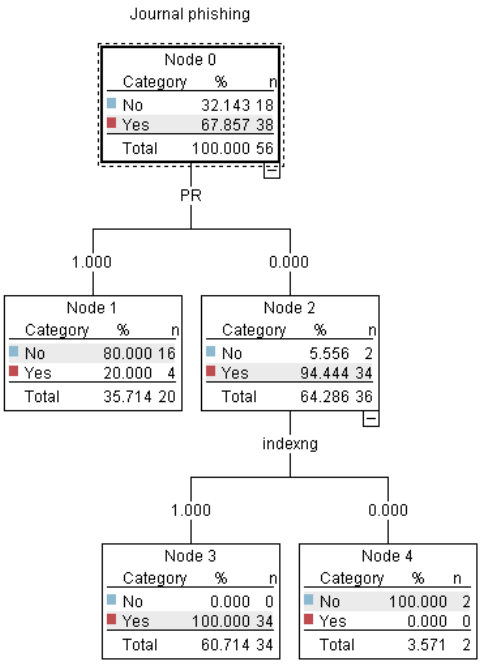
a) Domain lifetime as root feature



b) Availability of previous issue as a root feature.



c) Journal scope as root feature



d) Page ranking as root feature

Figure 3. Using of different features as root in decision tree

4. Measuring error ratio

To measure mentioned method error ratio, different dataset from previous training dataset should be used then according to achieved results, calculate error ratio. According to experiments, our approach resists against errors because if just one the root features be inaccessible, making decision is possible by choosing another feature as root.

5. Conclusion and Future Work

In this paper we discussed about a new kind of phishing attacks which were detected as journal phishing and mentioned our reason for this naming. Then, we detected key features of this kind of phishing attacks and presented a new approach for detecting them. Mentioned approach unlike all past methods to confronting phishing, has the ability to detect phishing journals. Using this method with the combination of past used methods to confront phishing attacks can be a part of future efforts. New presented features in this paper can be added to 27 known phishing attacks features and present a more perfect method rather than past methods to confronting to different kinds of phishing attacks.

Acknowledgements

We would like to acknowledge to Marwan M. Obeidat from Department of English Language and Literature, Hashemite University, Zarqa, Jordan.

References

[1] Dadkhah M, Jazi M.D, Lyashenko V. Prediction of phishing websites using classification algorithms based on weight of web pages characteristics. *Journal of Mathematics and Technology*. 2014; 5(2): 24-35. DOI: 10.7813/jmt.2014/5-2/4.

[2] San Martino A, Perramon X. Phishing Secrets: History, Effects, and Countermeasures. *International Journal of Network Security*. 2010; 11(3): 163-171.

[3] Mahmood M, Rajamani L. APD: ARM Deceptive Phishing Detector System Phishing Detection in Instant Messengers Using Data Mining Approach. *Global Trends in Computing and Communication Systems*. 2012; 269; 490-502.

[4] Li S, Schmitz R. *A Novel Anti-Phishing Framework Based on Honeypo*ts. eCrime Researchers Summit (IEEE). Tacoma, WA. 2009; 1-13. DOI:10.1109/ECRIME.2009.5342609.

[5] Johnson M. Eric. Managing Information Risk and the Economics of Security. 2009th Edition. Germany: Springer. 2009: 1-40.

[6] Dadkhah M, Davarpanah Jazi M. Secure Payment in E-commerce: Deal with Keyloggers and Phishings. *International Journal of Electronics Communication and Computer Engineering*. 2014; 5(3): 656-660.

[7] Alkhateeb F, Manasrah A, Bsoul A. Bank Web Sites Phishing Detection and Notification System Based on Semantic Web technologies. *International Journal of Security & Its Applications*. 2012; 6(4): 53-66.

[8] Schlegel R, Zhang K, Zhou X. Y, Intwala M, Kapadia A, Wang X. Soundcomber: A Stealthy and Context-Aware Sound Trojan for Smartphones. *In NDSS*. 2011; 11: 17-33.

[9] Chandavale A.A, Sapkal A.M. *Algorithm for Secured Online Authentication Using CAPTCHA*. Proceedings of the third International Conference on Emerging Trends in Engineering and Technology. Goa. 2010: 292 – 297. DOI: 10.1109/ICETET.2010.126.

[10] Hong J. The State of Phishing Attacks. *Communications of the ACM*. 2012; 55(1): 74-81.

[11] Agarwal N, Renfro S, Bejar A. *Yahoo Sign-In Seal and Current Anti-Phishing Solutions*. Proceedings of Web 2.0 Security & Privacy Workshop. 2007: 1-4.

[12] Aburrous M, Hossain M. A, Dahal K, Thabatah F. Intelligent Phishing Detection System for E-Banking Using Fuzzy Data Mining. *Expert Systems with Applications*. 2010; 37; 7913–7921.

[13] Shreeram V, Suban M, Shanthi P, Manjula K. *Anti-phishing detection of phishing attacks using genetic algorithm*. Proceedings of IEEE International Conference on Communication Control and Computing Technologies (ICCCCT). Ramanathapuram. 2010: 447 – 450. DOI: 10.1109/ICCCCT.2010.5670593.

[14] Chen J, Guo C. *Online Detection and Prevention of Phishing Attacks*. Proceedings of first International Conference on Communications and Networking (IEEE). China. 2006: 1-7. DOI: 10.1109/CHINACOM.2006.344718.

[15] Atighetchi M, Pal P. Attribute-based Prevention of Phishing Attacks. Proceedings of eighth International Symposium on Network Computing and Applications (IEEE). Cambridge. 2009: 266 – 269. DOI: 10.1109/NCA.2009.13.

[16] Dunlop M, Groat S, Shelly D. *Gold Phish: Using Images for Content-Based Phishing Analysis*. Proceedings of the fifth International Conference on Internet Monitoring and Protection (IEEE). Barcelona. 2010: 123–128.
[17] Mishra M, Gaurav, Jain A. A Preventive Anti-Phishing Technique using Code word. *International Journal of Computer Science and Information Technologies*. 2012; 3(3): 4248-4250.
[18] Liu G, Qiu B, Wenyin L. *Automatic Detection of Phishing Target from Phishing Webpage*. Proceedings of International Conference on Pattern Recognition (IEEE). Istanbul. 2010: 4153-4156. DOI: 10.1109/ICPR.2010.1010.
[19] Reddy V.P, Radha V, Jindal M. Client Side protection from Phishing attack. *International Journal of Advanced Engineering Sciences and Technologies*. 2011; 3(1): 39-45.
[20] Khonji M, Jones A, Iraqi Y.A. *Novel Phishing Classification Based On URL Features*. Proceedings of GCC Conference and Exhibition (IEEE). Dubai. 2011: 221 – 224. DOI: 10.1109/IEEEGCC.2011.5752505.
[21] Ruth Ramya K, Priyanka K, Anusha K, Jyosthna Devi CH, Siva Prasad Y.A. An Effective Strategy for Identifying Phishing Websites using Class-Based Approach. *International Journal of Scientific & Engineering Research*. 2011; 2(12): 1-7.
[22] M Jalalian, H Mahboobi. Hijacked Journals and Predatory Publishers: Is There a Need to Re-Think How to Assess the Quality of Academic Research?. *Walailak J Sci & Tech*. 2014; 11(5): 389-394.
[23] Dadkhah M, Obeidat MM, Jazi MD, Sutikno T, Riyadi MA. How Can We Identify Hijacked Journals?. *Bulletin of Electrical Engineering and Informatics*. 2015; 4(2): 83-87. DOI: 10.12928/eei.v4i2.449.
[24] Lukiæ Tin, Blešiæ Ivana, Basarin Biljana, Ivanoviæ Bibiæ Ljubica, Miloševiæ Dragan, Sakulski Dušan. Predatory and Fake Scientific Journals/Publishers– A Global Outbreak with Rising Trend: A Review.*Geographica Pannonica*. 2014; 18(3): 69-81.
[25] Williams E. Nwagwu. Open Access Initiatives in Africa- Structure, Incentives and Disincentives. *The Journal of Academic Librarianship*. 2013; 39(1): 3-10. DOI: 10.1016/j.acalib.2012.11.024
[26] R Kumar, R Verma. Classification Algorithms for Data Mining: A Survey. *International Journal of Innovations in Engineering and Technology*. 2012; 1(2): 7-14.