International Journal of Computer Applications Technology and Research
Volume 6–Issue 3, 133-140, 2017, ISSN:-2319–8656

Classification Model to Detect Malicious URL via Behaviour Analysis

N.Jayakanthan
Department of Computer Applications
Kumaraguru College of Technology
Coimbatore, India

A.V.Ramani
Department of Computer Science
Sri Ramakrishna Mission Vidyalaya College of Arts and Science
Coimbatore, India

Abstract: The challenging task in cyber space is to detect malicious URLs. The websites pointed by the malicious URLs injects malicious code into the client machine or steals the crucial information. As detecting a phishing URL is a challenging task, it is essential to enhance detection techniques against the emerging attacks. The most of the existing approaches are feature based and cannot detect dynamic attacks. Mostly the attacker uses the input form, active content and embeds @ symbol in URL for malicious attack. To detect this attack, a Behaviour based Malicious URL Finder (BMUF) algorithm is proposed. It analyzes the behaviour of the URL. The FSM based state transition diagram is used to model the URL behaviour into various states. The state transition from initial to final state is used for classification. This approach tests the genuine and malicious behavior of the URL based on the responses to the user. It accurately detects the nature of the URL.

Keywords: Malicious URL, Behaviour based Malicious URL Finder, Finite State Machine, Input form, Active Content

1. INTRODUCTION

The Malicious URL leads the user to phishing websites. These websites steal the user's confidential information without their knowledge using fake information form, active contents, and embed @ symbols in URL. These attacks inject malicious code in the client machine, and it controls the machine and spreads the malicious code to other machines in the same network[22]. The malicious web sites resemble the websites of the trusted organizations such as banks, government agencies, and e-commerce websites.

Generally most of the phishing attacks are Drive-by-downloaded attack. It installs the malicious code in users system to generate attack [7]. The code is automatically downloaded from the web page of the attacker without the permission of the user. This behavior is an important feature to detect web attacks.

The URL redirection mechanism is commonly used to carry out web attacks. The attacker redirects the visitor to the malicious website [15].The attacker r performs the following activities to make a successful attack. They are developing fraudulent websites and motivating users to visit those sites through malicious URL. The @ symbol is used to embed a malicious URL with a genuine URL. Apart from that input form, active contents also redirect the user to the malicious websites.

A number of approaches have been developed in recent years to detect the malicious attacks.

These include detecting suspicious websites [10], educating and training users [12], white list and black list based fault detection and feature based analysis of legitimate and malicious URLs.

Most of the web browsers are having built-in phishing detection abilities based on white and black lists. There exists no testing approach for anti-phishing professionals to manually verify suspected URL and intimate the administrators to take down the fake URLs. More over the phishers can exploit the cross site scripting (XSS) vulnerabilities by generating forms, active contents and @ symbol, motivating us to device behaviour based testing approach for malicious URL detection.

The proposed approach detects the malicious URLs based on the behaviour. Most of the existing approach detects the malicious URLs using lexical and host based features. But attacks in present scenario are highly dynamic which is not detectable through feature analysis. So we propose a behaviour based approach to detect the malicious URL.

The contribution of the proposed approach is as follows.

- It is a dynamic approach that detects the malicious URLs based on their behaviour.
- Behaviour based Malicious URL Finder algorithm is developed to detect the nature of the URL.
- FSM based state transition diagram is developed to model the URL behaviour in various states
- It improves the accuracy of the classification
- It is a light weight approaches capable of detecting malicious URL with low performance overhead.

The paper is organized as follows. Section 2 describes the related work done for malicious URL detection. Architecture of the proposed system is given in section 3. Section 4 deals with methodology. Section 5 discusses the analysis of the URL. Finally section 6 concludes the paper.

2. REVIEW OF RELATED WORK

Hossian Shahriar and Mohammed zulkernine [10] developed a tool phishTester to test the trustworthiness of the website

based on the behavior of the web application. They used Finite State Machine that captures the submission of forms with random inputs and corresponding responses.

Hyunsang Choi et al[11] analyze various types of discriminative features acquired from lexical, webpage, DNS, network, and link popularity properties of the associated URLs. The used SVM to detect malicious URLs, and both RAkEL and ML-kNN were used to identify attack types.

Sidharth Chhabra[6] et al and Y. Alshboul et al[1] found some malicious attacks obfuscating the host with largest host names, another domains and misspelled various. All these attacks hide the malicious URL behind the genuine URL. It leads the user to the malicious website.

Cheng Cao and James Caverlee[4] proposed a method to identify the malicious URLs through posting based features and click based features. The behavioral signals are analyzed for classification and this method yields 86% accuracy. Few machine learning approaches extract the URL features to train the classification model through training data. The features are categorized in to two classes- static and dynamic. In static analysis [4][13][14][2], the information is analyzed without the execution of the URL, but in dynamic approach the run time behavior of the URL is used for classification.

Charmi Patel and Hiteishi Diwanji[5] analyze the lexical and network based features using URL pattern matching algorithm. This algorithm analyzes the different patterns of URL to detect the malicious one. R.K. Nepali et al[16] use four machine learning algorithms - Naïve Bayes, random forest, support vector machine, and logistic regression to detect malicious URL.and obtain an accuracy of 97% using random forest algorithm. Y. Tao [21] proposed a dynamic method which mines the internet access log file to detect the malicious activity.

Peilin Zhao and Steven C H Hoi [18] proposed a Cost-Sensitive Online Active Learning (CSOAL) frame work to detect malicious URLs. The experimental results proved the efficiency of algorithm in classification. The black list based approaches [20][3][9] detects the URLs using the blacklisted profile. But they are incapable to detect emerging attacks.

H.K. Pao[17] et al method calculates Conditional Kolmogorov Complexity of the URL's with reference to genuine and malicious URLs. It compares the given URL with malicious or genuine URL databases for classification. W. Chu et[8] proposed a phishing detection method based on machine learning approach. The lexical and domain based features are analyzed for classification. This method properly classifies even the changes in the phishing URLs. E. Sorio[19] proposed a method to detect the hidden URLs based on their lexical features. Nearly 100 URLs are analyzed and experimental results show the efficiency of this approach

3.ARCHITECTURE OF THE PROPOSED SYSTEM

The architecture of the proposed system is given in figure 1. The components are browser, Behavioural Extraction, FSM Model, BMUL Classifier and Final classification.

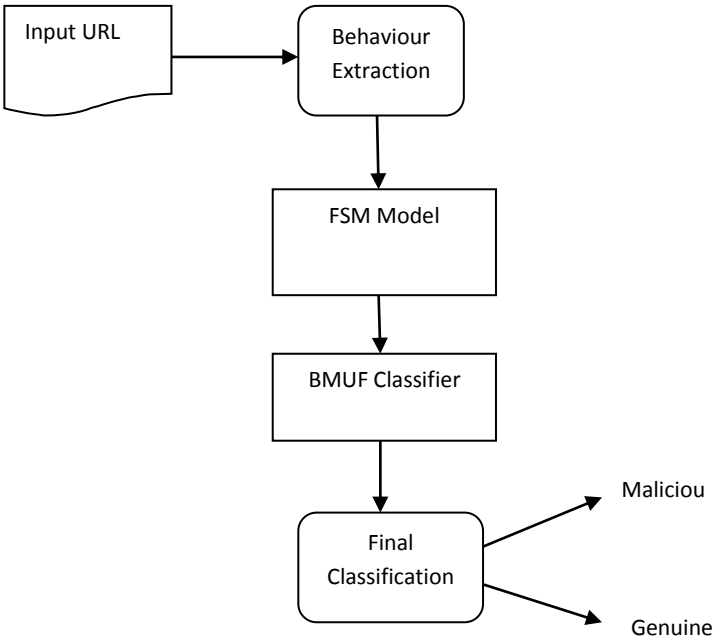

Figure 1.Architecture Diagram

available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

3.1Browser:

The URL is the input to the browser. The behaviour of the URL is extracted for the analysis.

3.3 FSM Model:

FSM state transition diagram model the URL behaviour in two various states. The states are derived from 3 inputs and 13 responses. The transition from the initial to final state leads to the classification.

3.3. BMUF Classifier:

The Malicious URL Finder (MUF) is rule based algorithm which analyzes the URL through using FSM state transition diagram. If any malicious behaviour is detected it marks the URL as malicious and collects the behaviour and reports it to the user.

4. METHODOLOGY

The proposed method uses the Behaviour based Malicious URL Finder (BMUF) algorithm to analyze the behavior of URL to detect whether the URL is genuine or malicious. The FSM based state transition diagram is used capture the URL's behavior in various states. The state transition from initial to finial state classifies the natural of the URL. This classifier improves the accuracy of the malicious URL Detection.

4.1 Algorithm:

Behaviour based Algorithm Malicious URL Finder (BMUF)

//Input: URL of the Webpage

//Output: Genuine or Malicious

M_B=Null // set of malicious behaviour.

Step1 : Consider the Input URL

If (Automatic content (ad) download occurs) then

Set Status = Malicious

$M_B = M_B$ U ad

Step 2 : Check the webpage pointed by the URL contains input form

When the user submits input form

a. If the user is redirected to a new webpage and URL of the page contain malicious words (mw) then

Set Status = Malicious

$M_B = M_B$ U mw

b. If the user is redirected to a new webpage and content is automatically downloaded (ad) then

Set Status = Malicious

$M_B = M_B$ U ad

Step 4 : Check the webpage pointed by the URL contains active content/s(ac) When the user access the active content then

a. If the user is redirected to a new webpage andURL of the page contain malicious words(mw) then

Set Status = Malicious

$M_B = M_B$ U mw

b. If the user is redirected to a new webpage and content is automatically downloaded (ad) then

Set Status = Malicious

$M_B = M_B$ U ad

Step 5: Check the webpage pointed by the URL contains @ symbol

a. It redirects the user to a webpage and URL of the webpage contain malicious word/s(mw) then

Set Status = Malicious

$M_B = M_B$ U mw

b. If the user is redirected to a new webpage and content is automatically downloaded (ad) then

Set Status = Malicious

$M_B = M_B$ U ad

Step 6 :

If status="Malicious" Then

Display URL is malicious

Display Set of Malicious behaviour M_S

Else

Display URL is genuine

4.2 FSM Model

The behaviours of the URL are modeled using Finite State Machine (FSM). Various symptoms of malicious and genuine URLs for FSM are developed based on submission of the information window, accessing active content. The norms are established by our literature survey. The malicious behaviours are identified as follows

a. Malicious content automatically downloaded from Web page of the URL

b. User access the input form or active content which leads to another webpage where content is automatically downloaded from the webpage.

c. Malicious Word or @ symbol in the URL.

The FSM is represented by $<Q, \sum, q_0, \delta, F>$ where F is the finite set of states, q_0 is the initial state, $\sum$ is the finite set of inputs, δ is the state transition function, and F is the set of final states.

(i) Q is a finite nonempty set of states. q_0 to q_{13} that represents the various behavioral states of URL.

(ii) $\sum$ is finite non empty set of inputs called an input alphabet. It is a combination of test cases $<I , K_i>$

(iii) δ is a function which maps $Q * \sum$ into Q and is usually called direct transition function. This is the function which describes the change of the state during transition. The mapping is usually represented by a transition table or transition diagram. The transition represents the behavioral change of the URL.

(iv) q_0 is the initial state. It represents the initial stage of the URL.

(v) F is the set of final states. It is assumed here that there may be more than one final state. The final states characterize the genuine or malicious behavior of the URL.

$\sum = <I_0,K_1>\ldots.<I_1,K_n>$ is the set of input symbols. Let q_0 be the initial state of the machine. It represents the input URL. The state q_1 of the machine for the input $<I_0,K_1>$ is as follows.

$q_1 = \delta(q_0 , < I_0,K_1>) = \delta 1(q_0 , < I_0,K_1>)$ where $\delta = \delta 1 : Q \times \sum$

The change in the state due to the second input symbol $< I_0,K_2>$) is q2.

$q_2 = \delta(q_1,< I_0,K_2>) = \delta(\delta 1(q_0, < I_0,K_1>),<I_0,K_2>)$

$= \delta_2 (q_0 , < I_0,K_1><I_0,K_2>)$ (1)

Where $\delta_2 : Q * I^2 \rightarrow F$

The function of the FSM is defined as follows

$q_n = \delta_n (q_0 , <I_0,K_1>.....<I_n,K_n>)$

$= \delta(\delta_{n-1} (q_0, , < I_0,K_1 >.....<I_{n-2} , k_{n-2} >,< I_{n-1} , k_{n-1} >)$ (2)

Equations 1 and 2 show the mapping function from one state to another state in the proposed approach

The FSM model is represented as a set of inputs (I_0 to I_2) and corresponding responses (K_0 to K_{13}) are discussed in detail in the following paragraphs. A URL is classified as malicious or benign based on the traversal from initial state to final state. The state transition is given in figure 2. The final stages (q_2, q_5, q_7, q_8, q_{11}, q_{13}) are legitimate and some of the final states (q_4, q_6, q_9, q_{10}, q_{12}) are phishing.

A state transition occurs for a given input and subsequent response. The transition is represented as <Input, response> pair as shown in the figure 2. The pair $<I_2, K_2>$ represents the input I_2 and its corresponding response K_2. There are three kinds of inputs.

1. The input URL (U),
2. URL leads to a webpage which contains malicious Active content.
3. URL leads to a webpage which contains input form

[Example:https://www.perspectiverisk.com/wp-content/uploads/2016/09/Login.png]

The features that represent the set of responses are given below.

iw : Indicates the user fills the information window and submits it.
ac : The user access the active content
@ : The presence of @ symbol in the URL
re : The page is redirected. It may happen due to the submission of input form or response of user interaction with active content or malicious domain pointed by the @ symbol.
mw : URL contains malicious word
p : Presence of the redirected page.
d : The content gets automatically downloaded from URLs web page. They are counterfeit executable programs.

These features are used to classify whether the URL is phishing or genuine. The ! symbol represents the absence of a particular feature (!iw represents the absence of the information window).

The proposed approach distinguishes the malicious URL from the legitimate one based on the behavior of the URL. The input and responses are given in the table1.

Table 1. Input and responses

Name	Representation	Explanation
I_0	U	Input URL
I_1	AC	The URL leads to a webpage that contains active content
I_2	I	The URL leads to a webpage that contains Input form.
K_1	!iw !ac !@ !re !mw !p !d	No Information window, no active content, no @ symbol present in URL, no redirection, no malicious word, no redirected page , no automatic content download.
K_2	iw !ac !@ !re !mw !p !d	User submit information window, no active content, no @ symbol present in URL, no redirection, no malicious word, no redirected page , no automatic content download.
K_3	iw !ac !@ re mw p !d	User submit information window , no active content, no @ symbol present in URL, redirection occurs, malicious word present in the URL, redirected page present, no automatic content download.
K_4	iw !ac !@ re !mw p !d	User submit information window , no active content, no @ symbol present in URL, redirection occurs, no malicious word, redirected page present, no automatic content

		download .
K_5	iw !ac !@ re !mw p d	User submit information window, no active content, no @ symbol present in URL, redirection occurs, no malicious word , redirected page present, automatic content download occurs.
K_6	!iw ac !@ !re !mw !p !d	No Information window present, active content occurs, no @ symbol present in URL, no redirection occurs, no malicious word, no redirected page present, no automatic content download
K_7	!iw ac !@ re !mw p !d	No Information window present, active content occurs, no @ symbol present in URL, redirection occurs, no malicious word , redirected page present, no automatic content download
K_8	!iw ac !@ re mw p !d	No Information window present, active content occurs, no @ symbol present in URL, redirection occurs, malicious word present in the URL, redirected page present, no content download occurs.
K_9	!iw ac !@ re !mw p d	No Information window present, no active content, no @ symbol present in URL, redirection occurs, no malicious word, redirected page present, automatic content download occurs. .

K_{10}	!iw !ac @ !re !mw !p !d	No Information window present, no active content, @ symbol present in URL, no redirection occurs, no malicious word, no redirected page present, no content download occurs. .
K_{11}	!iw !ac @ re mw p !d	No information window present, no active content, @ symbol present in URL, redirection occurs, malicious word present in the URL , redirected page present, no downloads occurs.
K_{12}	!iw !ac @ re !mw p !d	No information window present, no active content, @ symbol present in URL, redirection occurs, no malicious word, redirected page present, no downloads occurs.
K_{13}	!iw !ac @ re !mw p d	No information window present, no active content, @ symbol present in URL, redirection occurs, no malicious word, redirected page present, automatic content download occurs.
K_{14}	!iw ! ac !@ !re !mw !p d	No information window present, no active content, no @ symbol present in URL, no redirection occurs, no malicious word, no redirected page present, automatic content download occurs.

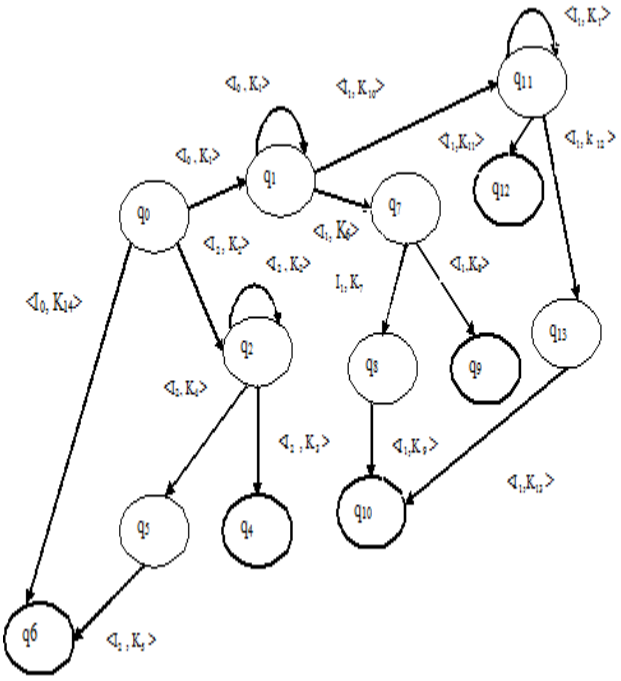

Figure 2 FSM state transition diagram

Figure 2 shows the state diagram of the FSM model. Here q_0 is the initial state. If the page downloaded from the given URL contains no information window, no active content, no @ symbol, no malicious word , no redirection and no automatic content downloads (code injection), then the next state is considered q_1. The finite set of states is as follows($q_1,q_2,q_3,q_4,,q_5,q_6,q_7,q_8,q_9,q_{10},q_{11},q_{12},q_{13}$). Each state represents the behavioral state of the URL.

4.3. State transition

The traversals of the URL help to identify it as malicious or legitimate. If the URL is having the state sequences given in table2, then it is considered as malicious (malicious final states are shown in a dark circle in the figure2.

Table 2 List of malicious states

States	Test cases	Description
q0 , q6	$<I_0,K_{14}>$	The content is automatically downloaded from the web page of the URL.
q0, q2, q4	($<I_2,K_2>$,$<I_2,K_3>$)	The user fills the information window and clicks the submit button, The user is redirected to a new page where the URL contains malicious word.
q0,q2,q5,q6	($<I_2$, $K_2>$, $<I_2$, $K_4>$, $<I_2$, K_5 >)	The user fills the information window and clicks the submit button, the user is redirected to a new page, and the contents are automatically downloaded from that page.
q0,q1,q7, q8,q10	($<I_0$, $K_1>$, $<I_1$, $K_6>$, < I_1, K_7 > , $<I_1,K_9>$)	The user accesses the active content, it leads the user to a new web page and the contents are automatically downloaded from the page.
q0,q1,q7,q9	($<I_0$, $K_1>$, $<I_1$, $K_6>$, $<I_1,K_8>$)	The user accesses the active content, it leads the user to a new web page where the URL contains malicious word
q0,q1, q11, q12	($<I_0$, $K_1>$, $<I_1$, $K_{10}>$, $<I_1,K_{11}>$)	The URL contain @ symbol it redirect the user to a another webpage it contains malicious word in the URL
q0,q1,q11,q13, q10	($<I_0$, $K_1>$, $<I_1$, $K_{10}>$, $<I_1$, K_{12} >, $<I_1,K_{13}>$)	The URL contains @ symbol that leads the user to a new web page and the contents are automatically downloaded from the page.

5.ANALYSIS OF THE URL

The set of URLs used for analysis are given below

1.https://www.perspectiverisk.com/wp-content/uploads/2016/09/Login.png

2.http://demo.smartscreen.msft.net/other/explotframe.html

3.www.yahoo.com

4.https://phishme.com/macro-based-anti-analysis/acf.css

5.http://geniune.com@malicious.com/config/change.html

The proposed Behaviour based Malicious URL Finder (BMUF) algorithm analyze the behaviour of each URL using various rules and classify it as genuine or malicious. The behavior and classification are given in the following table3.

Table 3 BMUF Classification

URL.NO	URL	Input Form	Active Content	@ Symbol	Redirection	Malicious Word	Auto Content download	Classification
1	Y	Y	N	N	Y	N	Y	Malicious
2	Y	N	Y	N	Y	Y	N	Malicious
3	Y	N	N	N	N	N	N	Genuine
4	Y	N	Y	N	N	N	Y	Malicious
5	Y	N	N	Y	N	Y	N	Malicious

Finite State Machine (FSM) model the behavior as various states. The state transition from the initial state to final state is used for classification. The state transition for each URL is given in table 4.

Table 4 State Transition

URLNO	States	Test case	Description	Classification
1	q0 , q6	$<I_0,K_{14}>$	The content is automatically downloaded from the web page of the URL.	Malicious
2	q0,q1,q7,q9	($<I_0$, $K_1>$, $<I_1$, $K_6>$, $<I_1,K_8>$)	The user accesses the active content, it leads the user to a new web page where the URL contains malicious word	Malicious
3	q0,q1	($<I_0$, $K_1>$)	URL is present but no malicious activities detected.	Genuine
4	q0,q1,q7, q8,q10	($<I_0$, $K_1>$, $<I_1$, $K_6>$, $<I_1$, $K_7>$, $<I_1,K_9>$)	The user accesses the active content, it leads the user to a new web page and the contents are automatically downloaded from the page.	Malicious
5	q0,q1, q11, q12	($<I_0$, $K_1>$, $<I_1$, $K_{10}>$, $<I_1,K_{11}>$)	The URL contain @ symbol it redirect the user to a another webpage it contains malicious word in the URL	Malicious

The classification of the list of URL is given in table 5.

Table 5 Classification

URL. No	URL	Classification
1	https://www.perspectiverisk.com/wp-content/uploads/2016/09/Login.png	Malicious
2	http://demo.smartscreen.msft.net/other/explotframe.html	Malicious
3	www.yahoo.com	Genuine
4	https://phishme.com/macro-based-anti-analysis/acf.css	Malicious
5	.http://geniune.com@malicious.com/config/change.html	Malicious

6. CONCLUSION:

The web attacks are challenging problem for the web users. Detecting malicious URL is a complex task due to the dynamic behavior of the URL. The proposed classification model to detect malicious URL is based on the behavior. The Behaviour based Malicious URL Finder (BMUF) algorithm analyzes the behavior in sequence of steps to detect the URL is genuine or malicious. Finite State Machine state transition diagram capture the behaviour into various states. The state transition from initial to final states leads to classification. Thirteen states are derived from 3 inputs and 13 responses. The final states represents whether a URL is genuine or malicious. The proposed algorithm improves the accuracy of the classification.

7. REFERENCES

1. Y. Alshboul, R. Nepali, and Y. Wang, “Detecting malicious short urls on twitter,” 2015.

2. Birhanu Eshete, A. Villafiorita, and K. Weldemariam, “Binspect: Holistic analysis and detection of malicious web pages,” in Security and Privacy in Communication Networks. Springer, 2013, pp. 149–166.

3. S. Bo, M. Akiyama, Y. Takeshi, and M. Hatada, "Automating url blacklist generation with similarity search approach," IEICE TRANSACTIONS on Information and Systems, vol. 99, no. 4, pp. 873–882, 2016.

4. C. Cao, J. Caverlee, Detecting spam urls in social media via behavioral analysis, in: Advances in Information Retrieval, Springer, 2015, pp. 703– 714.

5. Charmi Patel , Hiteishi Diwanji Research on Web Content Extraction and Noise Reduction through Text Density Using Malicious URL Pattern Detection",International Journal of Scientific Research in Science Engineering and Technology",Volume 2, Issue 3, ISSN : 2394-4099,May-June 2016.

6. S. Chhabra, A. Aggarwal, F. Benevenuto, and P. Kumaraguru, "Phi. sh/$ ocial: the phishing landscape through short urls," in Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference. ACM, 2011, pp. 92–101.

7. S.Chitra K. S. JayanthanS. PreethaR. N. Uma Shankar "Predicate based Algorithm for Malicious Web Page Detection using Genetic Fuzzy Systems and Support Vector Machine"International Journal of Computer Applications. Volume 40 - Number 10.2012. DOI:10.5120/5000-7277.

8. W. Chu, B. B. Zhu, F. Xue, X. Guan, and Z. Cai, "Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing urls," in Communications (ICC), 2013 IEEE International Conference on. IEEE, 2013, pp. 1990–1994.

9. M. Felegyhazi, C. Kreibich, and V. Paxson, "On the potential of proactive domain Blacklisting." LEET, vol. 10, pp. 6–6, 2010.

10. Hossain Shahriar and Mohammad Zulkernine Trustworthiness testing of phishing websites: A behavior model-based approach. Future Generation Comp. Syst. 28(8):1258-1271(2012).DOI 10.1016/j.future.2011.02.001.

11. Hyunsang Choi, Bin B. Zhu, Heejo Lee, "Detecting Malicious Web Links and Identifying Their Attack Types", InWebApps, June 2011.

12. D. Irani, S. Webb, J. Giffin, C. Pu, Evolutionary study of phishing, in: Proc. Of the 3rd Anti- Phishing Working Group eCrime Researchers Summit, Atlanta,Georgia, October 2008, pp.1–10.

13. J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: learning to detect malicious web sites from suspicious urls," in Proceedings of the 15th ACM international conference on Knowledge discovery and data mining. ACM, 2009, pp. 1245– 1254.

14.J.Ma, L. K. Saul, S. Savage, and G. M. Voelker "Learning to detect malicious urls," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, no. 3, p. 30, 2011.

15.Mitsuaki Akiyama, Takeshi Yagi, Takeshi Yada, Tatsuya Mori, Youki Kadobayashi,"Analyzing the ecosystem of malicious URL redirection through longitudinal observation from honeypots", Journal of Computer & Security, January 2017.

16.R. K. Nepali and Y. Wang, "You look suspicious!!: Leveraging visible attributes to classify malicious short urls on twitter," in 2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE, 2016, pp. 2648–2655.

17.H.K. Pao, Y.-L. Chou, and Y.-J. Lee, "Malicious url detection based on kolmogorov complexity estimation," in Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01. IEEE Computer Society,2012, pp. 380–387.

18.Peilin Zhao,Steven C H Hoi"Cost-sensitive online active learning with application to malicious URL detection",Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining,Chicago, Illinois, USA — August 11 - 14, 2013. 919-927.[

19. E. Sorio, A. Bartoli, and E. Medvet, "Detection of hidden fraudulent urls within trusted sites using lexical features," in Availability, Reliability and Security (ARES), 2013 Eighth International Conference on.IEEE, 2013, pp. 242–247.

20. B. Sun, M. Akiyama, T. Yagi, M. Hatada, and T. Mori, "Autoblg: Automatic url blacklist generator using search space expansion and filters," in 2015 IEEE Symposium on Computers and Communication (ISCC). IEEE, 2015, pp. 625–631.

21.Y. Tao, "Suspicious url and device detection by log mining," Ph.D. dissertation, Applied Sciences: School of Computing Science, 2014.

22. Zhi-Yong,Ran Tao, Zhen-He cai and Hao Zhang Li A. "Web Page Malicious Code Detect Approach Based on Script Execution". International Conference on Natural Computation 009.308-312. DOI:10.1109/ICNC.2009.363.