International Journal of Emerging Trend in Engineering and Basic Sciences (IJEEBS)
*ISSN (Online)*2349-6967
Volume 2 , Issue 1(Jan-Feb 2015), PP170-175

An Efficient Approach to Detecting Phishing Web

Mr. Chitte Bhushan Atmaram , Miss. Lande Sujata Chimaji ,

Mr. Nandiwale Basha Yallappa, Mr. Marathe Abhishek Vinod

Department of computer engg. , Govt College of engg, Awasari(Pune).

Abstract : As the Electronic Commerce and On-line Trade expand, phishing has already become one of the several forms of network crimes. This paper presents an automatic approach for intelligent phishing web detection based on learning from a large number of legitimate and phishing webs. As given a web, its Uniform Resource Locator(URL) features are first analyzed, and then classified by Naive Bayesian(NB)classifier. When the web's legality is still suspicious, its webpage is parsed into a document object model tree, and then classified by Support Vector Machine(SVM) classifier. Experimental results show that our approach can achieve the high detection accuracy, the lower detection time and performance with asmall sample of the classification model training set.

Keywords: Phishing; Naive Bays(NB); Support Vector Machine(SVM); Classifier

I. INTRODUCTION

1.1Problem Definition

This paper presents an automatic approach for intelligent phishing web detection based on learning from a large number of legitimate and phishing webs. Our approach determines whether a webpage is a phishing web or a legitimate one, based on its Uniform Resource Locator (URL) and webpage features, and is merely a combination of Naïve Bayes (NB) and Support Vector Machine (SVM).

1.2. Existing System

In existing system they were going to check URL with its feature. and then decided site is phishing or legitimate. And also check given URL in phishing and legitmate data if available then decided phishing or legitimate but it check only available data (URL) .

II.SYSTEM ARCHITECTURE:

In this section, a detailed discussion of our approach is provided to classifying webs reputation. The system architecture is shown in Fig. Our approach is performed in the following procedures.

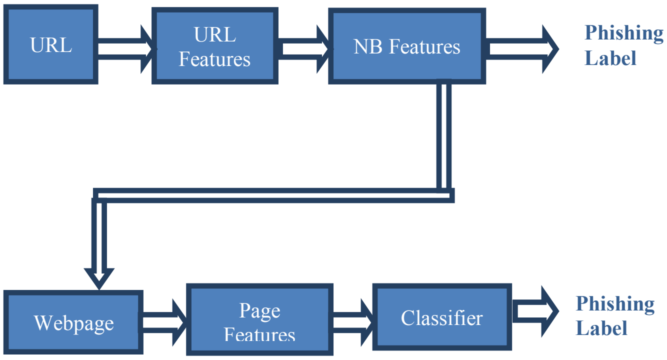


Fig 1. System Architecture

Given a web P, extract its URL identity and generate features.

1. Classify P by NB classifier and return result (+1: legitimate, -1: phishing or 0: suspicious).
2. If result=+1 or -1, output the phishing label. If result=0, go to Step 4.
3. If P has not a text input, output the phishing label as 1. If P has a text input, go to Step 5.
4. Extract its webpage identity and generate features.
5. Classify P by SVM classifier and output the phishing label.

URL Features Extraction:

-IP Address
-Number of Dots
-Suspicious URL
-Number of Slashes

NB Classifier:

-Each URL is represented by the above features (each having binary value)
-Probability is calculated if web belongs to either of two classes: Legitimate or Phishing, else Suspicious.

Webpage Feature Extraction:

-Forms:
Check for “input” 171labeled such as “password” or other personal data.
- Nil Anchors:
Check for anchors pointing to nowhere.
- Foreign Anchor:
Check for domain names in the current page and HREF tags
- Foreign Requests:
Check for requests to foreign domain names.
-SSL Certificate:
Check for SSL Certificate.

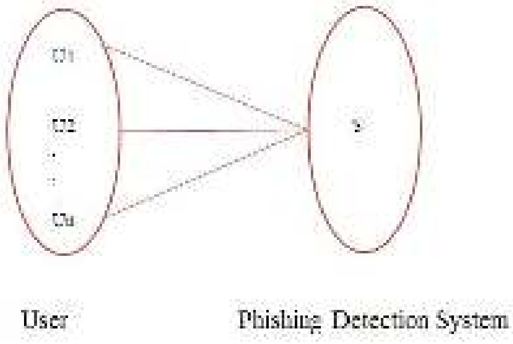
SVM Classifier:

-6-dimension feature vector is produced from the web feature extraction step.
-Least Squares Support Vector Machine (LS-SVM) is applied to this data for checking whether the web is legitimate.
-Probability is calculated if web belongs to either of two classes: Legitimate or Phishing.

III.MATHEMATICAL MODEL

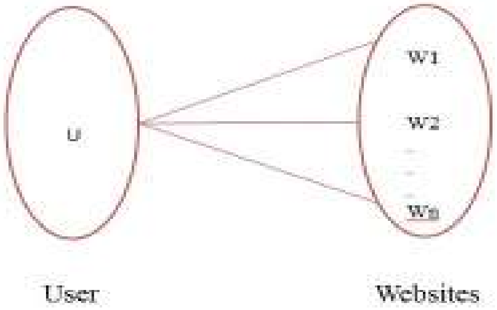
A] Mapping:

1]



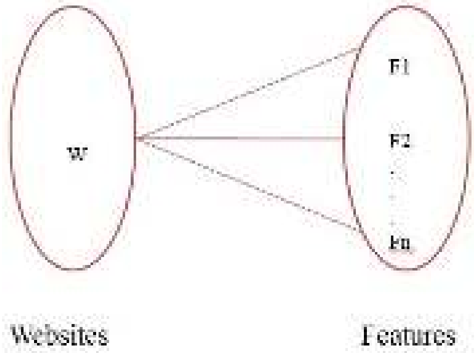
Many users can use one Phishing Detection System.

2]



One user can access many websites.

3]



One website can contain many features.

B] Set Theory

Our system can be represented as a set

$X = \{I ,O, S_C, F_C, C\}$

Where ,

I=set of inputs

O=set of outputs

S_C= set of outputs in success cases

F_C = set of outputs in failure cases

C = set of constraints

$I = \{W\}$

where,

W = set of URLs of websites

$O = \{R_P\}$

where,

R_P = set of results indicating whether website is phishing or not

$S_C= \{R_{Pn}\}$

where,

R_{Pn} = set of results correctly indicating whether website is phishing or not

$F_C = \{R_{Pn}, NULL\}$

where,

R_{Pn} = set of results wrongly indicating whether website is phishing or not

NULL representsno output

$C = \{C_1, C_2\}$

User

where,

Websites

C_1 = “User should enter a valid URL”

C_2 = “Phishing samples should be restricted to data collection obtained from PhishTank only”

$W = \{W_1, W_2,\ldots,W_n\}$

where,

W_1, W_2, …,$W_n$are URLs of websites.

R_P, R_{Pn}, $R_{Po}$are in the form

$R = \{R_1, R_2,\ldots,R_n\}$

where,

R_1, R_2,…,$R_n$are results indicating whether website is phishing or not

IV. CONCLUSION

In this paper, a novel approach is presented to identifying the potential phishing target of a given web. Every web claims a webpage identity, either real or fake. If a web claims a fake identity, abnormality may exist in a network space; therefore our approach could detect and differentiate a legitimate and a phishing web. Our approach first categorizes the URL features and test whether the page is phishing or not using NB. When the web's legality is still suspicious, then categorize its webpage features and test whether the page is phishing or not using SVM. The experimental results show that our approach has a high detection rate and a low false positive rate. In future works, the plan is to adjust existing feature extraction methods and seek for more relevant features to get a better result.

REFERENCES

[1] J. S. Downs, M. B. Holbrook, Decision strategies and susceptibility to phishing, in: Proc. The second symposium on usable privacy and security(SOUPS 2006), pp. 79-90.

[2] I. Bose, A.C.M. Leung, Unveiling the mask of phishing: threats, preventive measures and responsibilities, Communications of the Association for Information Systems 19 (24) (2007) 544-566.

[3]Google Inc, Google safe browsing for Firefox, http://www.google.com/tools/firefox/ safe browsing/

[4] Net craft Inc, Net craft antiphishing toolbar, http://toolbar.netcraft.com/

[5] Y. Pan, Anomaly based web phishing page detection, in: Proc. Twenty second annual computer security applications conference(ACSAC'06), 2006, pp. 381-392.

[6] I. He, S.J. Horng , An efficient phishing webpage detector, Expert Systems with Applications 38 (2011) 12018-12027.

[7] X. Chen, I. Bose, Assessing the severity of phishing attacks: A hybrid data mining approach, Decision Support Systems 50 (2011) 662-672.

[8] H.Wang , B.Zhu, C.WANG, A Method of Detecting Phishing Web Pages Based on Feature Vectors Matching, Journal of Information and Computational Systems 2012 Vol. 9 (15): 4229-4235.

[9] W.Zhuang , Q. Jiang, Intelligent Anti-phishing Framework Using Multiple Classifiers Combination, Journal of Computational Information Systems 2012 Vol. 8 (17): 7267-7281.

[10] Y. Zhang, J.Hong, CANTINA: A content-based approach to detecting phishing web sites, in: Proc. the international World Wide Web conference(WWW), 2007, pp. 639-648.

[11] D. K. McGrath, M. Gupta, Behind Phishing: An Examination of Phisher Modi Operandi, in: Proc. the USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET), 2008, pp.1123-1136.