ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 4, Issue 4, April 2017

Security Evaluation of Pattern Classifier against Phishing URL Detection

[1]Miss J. N. Karnase, [2]Dr. V. M. Thakare, [3]Dr. S.S.Sherekar
[1] [2] [3]SGBAU, Amravati, Maharashtra, India.
[1]jyoti.karnase11@gmail.com, [2]vilthakre@yahoo.co.in, [3]ss_sherekar@rediffmail.com

Abstract :- — **Pattern classification is a branch of machine learning that focuses on recognition of patterns and regularities in data. In adversarial applications like biometric authentication, spam filtering, network intrusion detection the pattern classification systems are used. Extending pattern classification theory and design methods to adversarial environment is thus a novel and very relevant research direction. Spam filtering to discriminate between a "legitimate" and a "malicious" pattern class. Intrusion analysis is the process of combing through IDS alerts and audit logs to identify real successful and attempted attacks. Phishing is a social engineering attack that exploits user's ignorance during system processing has an impact on commercial and banking sectors. Numerous techniques are developed in the last years to detect phishing attacks such as authentication, security toolbars, blacklists, phishing emails, phishing websites, and URL analysis, In this paper, present phishing detection system using features extracted from URLs lexical only to meet two important goals which are wide scope of protection and applicability in a real-time system and calculate the probability of characters sequence in URLs using the N-gram model.**

Keywords: **- Adversarial Classification, Machine Learning, Security Evaluation, Feature Extraction, Malicious URL Detection.**

I. INTRODUCTION

Pattern classification is a branch of machine learning that focuses on recognition of patterns and regularities in data. In adversarial applications like biometric authentication, spam filtering, network intrusion detection the pattern classification systems are used. Several works have addressed the problem of designing robust classifiers against these threats, although mainly focusing on specific applications and kinds of attacks. [1]. Email plays an increasing important role in our daily life. Spam filtering to discriminate between a "legitimate" and a "malicious" pattern class. Pattern classification systems may exhibit vulnerabilities, whose exploitation may severely affect their performance, and consequently limit their practical utility,. Previous work has been mainly focused on devising adversary aware classification algorithms to counter evasion attempts, application of feature selection. Multimodal biometric systems are commonly believed to be more robust to spoofing attacks than unimodal systems, as they combine information coming from different biometric traits. this problem is the high false-positive rate in the sensors used by IDS systems to detect malicious activities. Mainly three main open issues can be identified: (i) analyzing the vulnerabilities of classification algorithms, and the corresponding attacks; (ii) developing novel methods to assess classifier security against these attacks, which is not possible using classical performance evaluation methods (iii) developing novel design methods to guarantee classifier security in adversarial environments [5].

In this paper, present phishing detection system using features extracted from URLs lexical only to meet two important goals which are wide scope of protection and applicability in a real-time system and calculate the probability of characters sequence in URLs using the N-gram model, this method based on the probability of a gram with the assumption that the probability of any gram just depends on the previous gram probability and calculate the probability of characters sequence in URLs using the N-gram model.

II. BACKGROUND

In adversarial classification tasks like spam filtering, intrusion detection in computer networks, and biometric identity verification, malicious adversaries can design attacks which exploit vulnerabilities of machine learning algorithms to evade detection, or to force a classification system to generate many false alarms, making it useless. Several works have addressed the problem of designing robust classifiers

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 4, Issue 4, April 2017

against these threats, although mainly focusing on specific applications and kinds of attacks. [1].

Multimodal biometric systems are commonly believed to be more robust to spoofing attacks than unimodal systems, as they combine information coming from different biometric traits. Recent work has shown that multimodal systems can be misled by an impostor even by spoofing only one biometric trait. This result was obtained under a 'worst-case' scenario. Attacks, and to design robust fusion rules, without the need of actually fabricating spoofing attacks [2].

Email plays an increasing important role in our daily life. The series spam problem causes a huge economic lost. The spam is defined as the unsolicited commercial email. To avoid the detection of spam filter, the spammers change the behaviors of junk mails to mislead the decision of the classifier. This problem is called the adversary learning, which means the spammer will intentionally modify the samples to confuse the spam filter. The adversary attack can be classified into exploratory and causative attack [3].

Intrusion analysis, i.e., the process of combing through IDS alerts and audit logs to identify real successful and attempted attacks, remains a difficult problem in practical network security defense. So that analysts' time can be saved [4].

Pattern classification systems are commonly used in adversarial applications, like biometric authentication, network intrusion detection, and spam filtering. The input data can be purposely manipulated by an adversary to make classifier to produce false negative. As this adversarial scenario is not taken into account by classical design methods, pattern classification systems may exhibit vulnerabilities, whose exploitation may severely affect their performance, and consequently limit their practical utility [5].

This paper present brief introduction of Phishing URL detection in **Section I**. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses Existing Methodologies. **Section V** discusses proposed methodology. **Section VI** Advantages & Disadvantages. Finally **section VII** Conclude this review paper.

III. PREVIOUS WORK DONE

In the research literature, Battista B et al.(2011)[1] The generative model of data distribution for adversarial classification problems, which takes explicitly into account the presence of a malicious adversary based on such model then proposed a method for robust classifier design, for generative classifiers.

B. Biggio et al.(2012)[2] used to investigated the robustness of different score fusion rules for multimodal biometric verification systems, against spoofing attacks and focused on a bimodal system consisting of fingerprint and face biometrics. A large number of data sets including real spoofing attacks.

Junyan Peng et al.(2013)[3] the method to tackle the focus attack in spam filter instead of eliminating the attack samples, the proposed method is to reduce the effect of focus attack. For each feature in Naive Bayes classifier .

L Zomlot et al.(2013)[4] the application of machine learning has a different objective. Instead of using machine learning to make a decision on whether an event is malicious or not, and use it to prioritize alert correlation graphs from an upstream analysis tool..

Battista Biggio et al.(2014)[5] the main contribution is a framework for the empirical evaluation of classifier security evaluation, based on the definition of potential attack scenarios. Authors proposed: (i) a model of the adversary (ii) a corresponding model of the data distribution; and (iii) a method for generating training and testing sets that are representative of the data distribution, and are used for empirical performance evaluation.

IV. EXISTING METHODOLOGIES

A. The model of data distribution for adversarial classification

A model of data distribution in presence of attacks, and show how it naturally suggests a general method for robust design of generative classifiers model of data distribution in adversarial environments introduce a Boolean random variable $A \in \{T, F\}$ which determines whether the sample being generated is subject to the attack ($A = T$) or not ($A = F$). Exploiting the model for designing robust classifiers the corresponding Bayesian network used [1].

B. Multimodal biometric systems

Multimodal biometric systems have been originally proposed to improve the personal identity recognition performance only consider fusion at the matching score level, it is the most commonly adopted without loss of generality also focus on a system made up of a fingerprint and face matcher finally the matching scores are combined through a fusion rule which outputs a new real-valued score $f(s_1, s_2)$: the claimed identity is accepted and the person is classified as a genuine user, if $f(s_1, s_2) \geq s^*$; otherwise, it is classified as an impostor. [2].

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 4, Issue 4, April 2017

C. The revised Naive Bayes classifier

A classifier model based on the Naive Bayes classifier to confront the adversary attack, The Naive Bayes classifier is selected due to its low time complexity The purpose of ri is to lower the significance of the weight Wi when the i^{th} feature Ti appears both in spams and hams with the similar probabilities, which is suspicious and likely to be attacked.The Naive Bayes classifier, where R = (r1 r2, • • • ,r_p).ri of the ith feature Ti is calculated

ri= $\frac{n(Ti \mid Ds)}{n(Ds)}$ $\frac{n(Ti \mid Dh)}{n(dh)}$ Where n(Ti |Ds) and n(Ti | Dh) are the number of feature T occurs in spams and hams and, n(Ds) and n(D_h) are the total number of spam emails and ham emails [3].

D. Classify the correlation graphs from SnIPS, Feature selection, Learning Approaches

A classifier model based on the Naive Bayes classifier to confront the adversary attack the Naive Bayes classifier is selected due to its low time complexity. To classify the correlation graphs from SnIPS into the following two classes Interesting and Non interesting, Learning Approaches Supervised and semi supervised Learning.[4].

E. a model of the adversary, a model of the data distribution, a method for generating training and testing sets

A model of the adversary, that allows us to define any attack scenario, a model of the data distribution, that can formally characterize this behavior p(X | Y , A=F) = pD(X | Y), According to this equation, it can be defined as the empirical distribution of D The above generative model of the training and testing distributions ptr and pts is represented by the Bayesian network. a method for generating training and testing sets that are representative of the data distribution, and are used for empirical performance evaluation [5].

V. ANALYSIS AND DISCUSSION

In many applications, the design of robust classifiers against exploratory integrity attacks. In general, to counteract these attacks, a classifier has to be learnt on an hypothesis distribution $P_{ts}(X, Y)$, trying to prevent unknown attacks. For simplicity, in this work consider *generative* classifiers, as they can be directly learnt on the assumed $P_{ts}(X, Y)$ According to our model, to define $P_{ts}(X, Y)$, one has to set the probability distributions $P_{ts}(A = T)$, $P_{ts}(Y \mid A = T)$ and $P_{ts}(X \mid Y, A = T)$. The distributions $P_{ts}(Y \mid A = F)$ and $P_{ts}(X, Y \mid A = F)$ are instead identical to the corresponding distributions of the data *D* collected for classifier training, and can thus be estimated from *D* [1].

A multimodal system operates as at the design phase, authorized users (clients) are enrolled their biometric traits are stored in a database, together with the corresponding identities. Finally, the matching scores are combined through a fusion rule which outputs a new real-valued score f (s1, s2): the claimed identity is accepted and the Person is classified as a genuine user, if f(s1,s2) ≥s*; otherwise, it is classified as an imposter. The term s* is an acceptance threshold. Score-level fusion rules can be subdivided into fixed and trained. The difference between them is that the latter include a set of parameters to be estimated from training data [2].

Revised Naive bayes classifier method is more robust when the attack degree increases ri is to lower the significance of the weight Wi when the i^{th} feature Ti appears both in spams and hams with the similar probabilities, which is suspicious and likely to be attacked. It is obvious that the value of ri ranges from 0 to 1. When Ti exists both in spams and hams with equal probabilities, ri approaches 0 making the importance of T; lower. When T; exists mainly in spams or hams, ri approaches 1 and the revised Naive Bayes classifier becomes the original one. The accuracy on the attacked samples of the proposed method is higher than standard Naive Bayes classifier [3].

Conducted experiments with normalized polynomial, Gaussian Radial Basis Function(RBF) kernel, and Pearson Universal Kernel (PUK)is a universal kernel that can be calibrated to work as any of the standard SMO kernels [4].

Use a publicly available email corpus, TREC 2007, Where TR and TS represents training and testing set respectively. It consists of 25,220 legitimate and 50,199 real spam emails. The features or words are removed from TR using the SpamAssassin tokenization method, the classifier SVMs are implemented with the LibSVM; Logistic Regression (LR) is used for practical analysis [5].

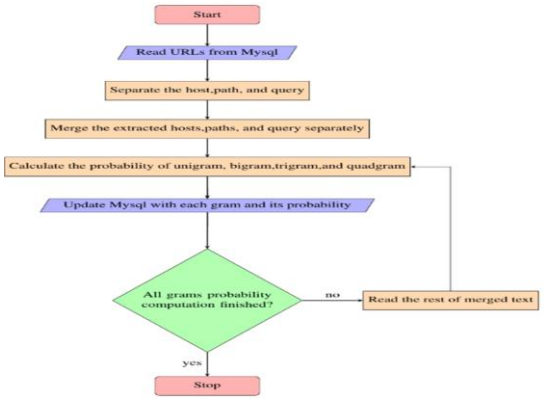
VI. PROPOSED METHODOLOG

Malicious Url Detection

Phishing is a social engineering attack that exploits user's ignorance during system processing has an impact on commercial and banking sectors. Numerous techniques are developed in the last years to detect phishing attacks such as authentication, security toolbars, blacklists, phishing emails, phishing websites, and URL analysis. Regrettably, nowadays detection system implemented for specific attack vectors such as email which make developing wide scope detection is

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 4, Issue 4, April 2017

much needed. Previous studies show that analysis of URLs proved to be a good option to detect malicious



activities where this method mostly based on features of lexical, host information, and other complex method which requires a long processing time. In this paper, present phishing detection system using features extracted from URLs lexical only to meet two important goals which are wide scope of protection and applicability in a real-time system.

Modules

1) Dataset collection

Collect phish websites from Phish tank which have been accumulated over a long period. We propose a novel method to divide dataset into three different datasets according to the years they appear in the dataset.

2) Dataset Processing

Data pre-processing is to get an understandable format from the collected datasets (raw data) because of inconsistently, incompleteness, and certain behaviors lacking among the main features of real life data. To fix such issues, data pre-processing is implemented to make the raw data format suitable for further processing.

3) Individual Classifier Evaluation

Classification is predicting the class label of input samples. There are two outputs in a binary classification problem such in phishing detection in this paper the output is either "1" or "0". Several methods can be used to measure classification performance where the most popular metrics are accuracy.

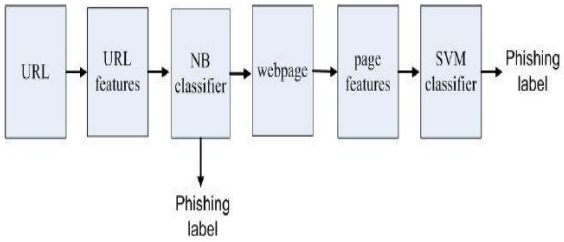
Algorithm

N-Gram Algorithm

N-grams of texts are extensively used in text mining and natural language processing tasks. They are basically a set of co-occurring words within a given window and when computing the n-grams you typically move one word forward .Parsing N-sized grams from the training data set is a general method to generate the language models where N is an integer. Markov chains are used to construct the N-grams where this method based on the probability of a gram with the assumption that the probability of any gram just depends on the previous gram probability and calculate the probability of characters sequence in URLs using the N-gram model whereby a lot of works proved that the N-gram method is the best option to generate the language model and tough to improve on it. In spite of the N-gram method to build the URL language model is implemented in previous work and used for malicious websites detection, to our knowledge this technique is not used for pure phishing URL detection yet. This work tries to implement unigram, bigrams, trigrams and four grams. Proposed methodology to create our N-gram models based on benign dataset only.

System architecture

In this section, a detailed discussion of approach is provided to classifying webs reputation. The system architecture is shown in Fig. 1. The approach is performed in the following procedures:



Step 1 Given a web P, extract its URL identity and generate features.
Step 2 Classify P by NB classifier and return result (+1, -1 or 0).
//+1: legitimate, -1: phishing, 0: suspicious
Step 3 If result=+1 or -1, output the phishing label,
If result=0, go to Step 4.
Step 4 If P has not a text input, output the phishing label (1).
If P has a text input, go to Step 5.
Step 5 Extract its webpage identity and generate features.
Step 6 Classify P by SVM classifier and output the phishing label

VI. POSSIBLE OUTCOME AND RESULT

Detect phishing attacks such as authentication, security toolbars, blacklists, phishing emails, phishing websites, and URL analysis.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 4, Issue 4, April 2017

Nowadays detection system implemented for specific attack vectors such as email which make developing wide scope detection is much needed. Phishing detection system using features extracted from URLs lexical only to meet two important goals which are wide scope of protection and applicability in a real-time system. Calculate the probability of characters sequence in URLs using the N-gram model whereby a lot of works proved that the N-gram method is the best option to generate the language model and tough to improve on it. N-gram based features provide high accuracies

VII. CONCLUSION

Phishing is a social engineering attack that exploits user's ignorance during system processing has an impact on commercial and banking sectors. Numerous techniques are developed in the last years to detect phishing attacks such as authentication, security toolbars, blacklists, phishing emails, phishing websites, and URL analysis. The wide scope of phishing detection classifier is presented in this paper. The accuracy level is achieved without huge time consuming as in previous works. Although, N-gram based features provide high accuracies

VIII. FUTURE SCOPE

There is still a gap to further improve the accuracies and reduce the error rate and believe in adding the most effective and lightweight bag of word features will improve the accuracy and reduce the error rates rapidly.

REFERENCES

[1] Battista Biggio, Giorgio Fumera, Fabio Roli "Design of Robust Classifiers for Adversarial Environments" IEEE , Cagliari, Italy ,p 977-982, 2011.
[2] Biggio Z. Akhtar G. Fumera G.L. Marcialis F. Roli "Security evaluation of biometric authentication systems under real spoofing attacks" IET Biometrics ,vol 1,No 1 ,PP.11-24, February 2012.
[3] JUNYAN PENG, PATRICK P. K. CHAN "Revised Naive Bayes Classifier For Combating The Focus Attack In Spam Filtering." International Conference on Machine Learning and Cybernetics, Tianjin, P 610-614, July 2013.
[4] Loai Zomlot , Sathya Chandran , Doina Caragea , Xinming Ou "Aiding Intrusion Analysis Using Machine Learning" IEEE , PP.40-47. 2013
[5] Battista Biggio, Giorgio Fumera, Fabio Roli "Security Evaluation of Pattern Classifiers under Attack", IEEE transaction on knowledge and data engineering, vol 26 ,No 4, April 2014