ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Exploration of the Semi-Supervised Learning Approach for Detecting Phishing Attacks.

Vignesh Jeevan*, Refonaa J, and Suraj Shaurya.

Computer Science and Engineering, Sathyabama University, Chennai, Tamil Nadu, India.

ABSTRACT

This paper proposes the phishing attacks that can be done in the web page. We are using the transductive algorithm to prevent phishing attacks. An email filtering technique is used to filter the mails from the spam mails. The spam mails which are sent by the hackers are the malicious mails. Most of the mail servers could not able to detect those spam (malicious) mails. We also use the password shuffling technique to improve protection. The password shuffling allows the user the shuffle the password n number of times. The captcha is used to verify whether the user is a human and not robot. In this paper we discuss about the motion graphical captcha. In this paper we also detect the targeted spam mails which is send from the local network.

Keywords: Motion Captcha, Password Shuffle, Classifier.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

In this document we are going to discuss about the various phishing activities which are done in day-today-activities.

Nowadays email has become the most important tool of the human. It is very important to protect the email from phishing attacks. To protect the email we will be using various algorithms including the TSVM-Transductive Support Vector Machine algorithm.

The email is an important approach for every people in this world. In the email system we have a client (user) and a mail server which handles emails and stores all the emails. There are various approach to send the email like Post office Protocols (POP), Interactive Mail Access Protocols (IMAP), POP version 3 (POP3). POP is used to send and receive the messages offline. It is mainly used for single user computers. In this POP the mails are directly delivered to the users inbox. In which the user can access the emails. The messages which are delivered those mails used to get deleted from the server. The POP is supported by any email client like Yahoo, Gmail, Outlook etc.

IMAP is another type of email protocol which is used to send and receive the emails when user reading the current emails. In the IMAP the emails are directly delivered to the server in which the user has to connect to the server(email client) and read the emails. The IMAP also offers many advantages like if one email is read in one computer if the user logs into another computer in doesn't the read email as unread. Webmail is similar to the IMAP. In webmail all the emails are downloaded to the user's computer. In which those emails can be read offline.

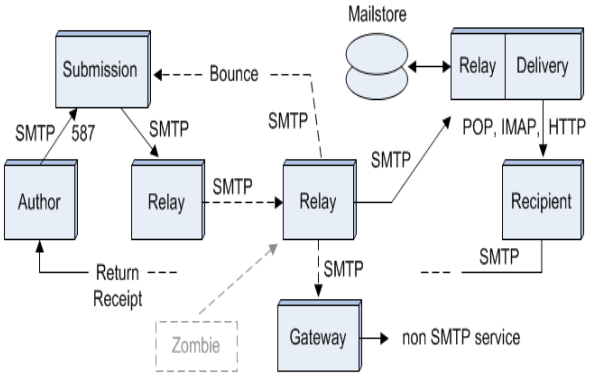

Fig.1. Process of delivery of email.

In the above diagram the process how the email is delivered is shown in detail. In this we use a relay concept to relay the emails. In this diagram we use a Simple Message Transfer Protocol (SMTP). As the message is delivered it goes to the SMTP server which the relays to the appropriate reciepeint through the address.

Existing System

In the current system of email we have a username and a password. If the user enters the correct username and the password the user is allowed to view his inbox. The main drawback of this system is it is very easy for the hacker to hack the email. Also in the current system we have an approach if we forgot our password. In this approach the user is given an OTP password sent to the registered phone number. If the OTP matches the user is allowed to set a new password for his email.

We also have captcha which is not used by most of the email servers. In this captcha the user has to type the given text to prove that the user is not a robot or a hacker. In the current system there are two types of captcha. They are numerical captcha and the image captcha.

ISSN: 0975-8585

The main drawback of the numerical and image captcha is that it can be easily hacked by the hacker. These captcha does not improve the protection of the email.

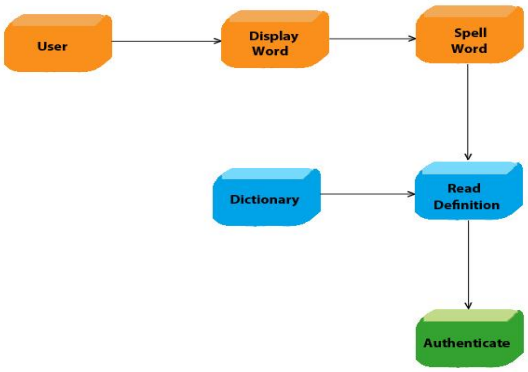

Fig.2. Block diagram of Numerical Captcha.

In the above diagram the block diagram of numerical captcha is shown in detail. The process starts with the user where the words would be displayed to the user. The words would be from the appropriate dictionary. After the user enters the word the system authenticates the word. If the word is correct the user is allowed to access his email.

The image captcha is similar to the numerical captcha where the user is given a image. The image could be of a number or some places where the user has to identify or he has to enter the numbers from the image.

If the entered image is correct the user is allowed to view his email. The main drawback of the image captcha is that sometimes the user does not able to understand the image given by the system. This is because the image is not properly taken of the result of image which is not given properly by the system.

Server-based scheme, such as shared e-certificate and dynamic security skins. They allow the remote server to provide the unique identity to the client to verify whether it matches or not. Browser also has the anti-phishing schemes for filtering the phishing web pages. Example- Internet Explorer, Google Chrome, Firefox. The requested webpage would be checked in the phishing filter where there would be the list of harmful websites. The average time the phishing website stays online is about 3.8 days. The longest is only 3.8 days.

PROPOSED SYSTEM

We proposed an approach for detecting the phishing web pages using various methods. We are going to use the Transductive Sector Vector Machine algorithm (TSVM). This algorithm is mostly used in webpages to improve the protection.

We are going to detect the phishing of web page with the web image and DOM objects. First we will extract the features from the web page using the TSVM algorithm. We are going to include the shuffling of passwords. The shuffling of passwords is used to improve the strength of the password. In this approach we can shuffle the passwords n number of times. Here we are taking the password as the string rather than character.

Motion graphical captcha is an approach to improve the security of the email. In this type of captcha the user is allowed to move the captcha object with the help of the mouse cursor. For this we are going to use the pointer technique.

We are also including the classification approach which is used to classify the spam mails from our inbox. The classification approach is the most important approach to improve the protection of email. As now

a days we are getting malicious mails from the unknown source where the email server could not detect those mails as spam. So we use this approach so those mails could be detected and can be removed.

For doing this classification we have to retrieve the mail data from the email server. As we are retrieving the mails from the mail server and then applying the algorithm to detect the spam mails.

Modules

GUI Design

This module is used to design the main user interface of the password shuffler. The GUI should contain the username and the password where the user should be able to type his username and his password. The GUI should be made easy so that the user should not have any problem while logging the system.

The design could be made with the textbox and the password box. When the user enters the password the password should be encrypted in the stars or in the dots.

Filtering based of the keywords

Spam filters are used to filter the mail inbox from the spam mail and also used to detect which are the spam mails in our inbox. With this type of filtering the user could be able to easily detect the spam from his email and could be able to delete those mails.

Spam filtering is mainly used to block the malicious mails from entering the inbox but it will land directly into the spam box. Spam box is the place where there will be all the spam mails of the user. In the below diagram we show how the mails are classified as the spam.

For doing the spam filter we are going to check the keywords using single or multiple keyword. The process starts with the email which arrive to the email address of the user. Then we apply the degree of membership value and we compare those with the membership categories. Fuzzy similarity is an approach to calculate the malicious words in the mail. If the words get detected it gets classified as the spam. If the mail is not a spam it gets legitimate and gets stored in the mail box else it gets quarantine.

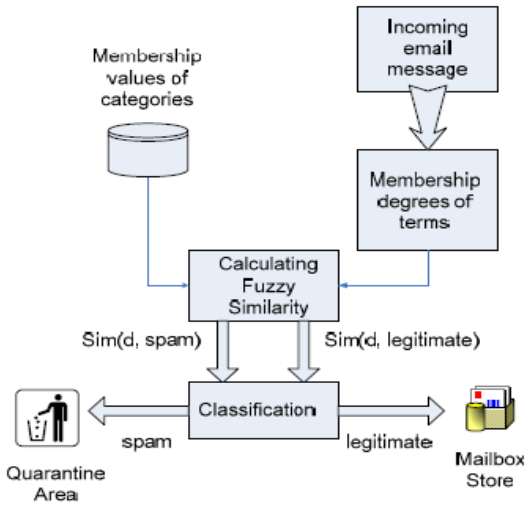

Fig.3. Block diagram of Password similarities.

ISSN: 0975-8585

Classification

In this module we are going to synchronize all the mails from the mail server. Classification is based on calculating the fuzzy similarity of the received mails. In order to calculate the fuzzy similarity we have to determine the membership degree of each token to the message.

With this we could be able to detect the spam mails. It will classify the mail according to the mail no, sent date, description and spam. To detect the spam we need to determine the frequency of each token in the message. The token with the maximum number of occurrences will be assigned a value of and all other tokens will be assigned proportional values.

The below diagram shows how the mails are processed. The email goes through the pre processing where the mail is checked with the company directory and with the data sets. The processed mails goes for extraction using random forest. After extraction the mail is classified either targeted mails or non-targeted mails.

The main advantage of the classification is the user can easily identify the spam mails from his email id. The user is allowed to delete the malicious mails from his inbox. The malicious mail could also contain the attachments which could harm the user system. So this system will also scan the attachments of every email and give the result to the user.

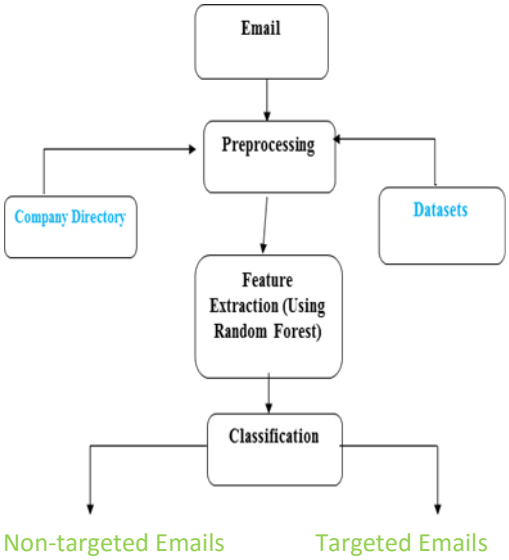

Fig.4. Block diagram of Classification.

Password Shuffling

The most common type of attack is the guessing of passwords. Any attacker can guess password locally or remotely using either a manual or automated approach. Password guessing is never difficult. As any known person of us can easily guess password with the hints of the family or favourite colour or teams.

The password could be easily guessed by the person who sits near you. For this we are going to shuffle the password n times where the user could be able to shuffle his password every time he types his password.

In this approach we are going to take the password as the string. In the existing system we used to take the password as character. As we are taking as string the single words of the password is taken as the input. This allows the user to shuffle his password. But this method won't work if we are taking the password as the character.

The main drawback of the password shuffling is as we are increasing the attempts of the password which could be profit for the hacker to easily hack the email. The hacker could be able to guess the password as we have shuffled the password.

Motion Captcha

The existing system has two types of captcha they are numerical captcha and image captcha. The drawback of the existing system is that they are easily hacked by the attacker. For improving the security we are using motion captcha.

The motion graphical captcha is used to validate whether the user is a human and not robot. For this we will be designing the user interface of the captcha. The captcha will be having the object. That object can be moved through the window using the cursor. The captcha will be validated with the pre combinations given by the system.

The user has to solve the captcha with the given combinations. If the user solves the captcha correctly the user could access the inbox else the user will be given another combination. There will be a limit of attempts up to two. If the user exceeds the limit then the server will generate a OTP (One Time Password) and send it to the registered phone number. The user has to enter the OTP to access his inbox.

The main advantage of the motion captcha is that it improves the protection of the email. The motion captcha can be implemented in various mail servers. This captcha uses the swing concept. The swing concept in java is used to move the object. The object has to be moved to the correct place.

The combinations will be generated by the server every time the user wants to access the email. These combination will never be repeated. The combinations will be in the graphical format where it is very difficult for the hacker to attack.

REFERENCES

[1] Arjun shah, varun das. IJAR 2015; 10 (3) : 23-88.
[2] Chandrasekaran M, Shakaranarayanan V. IJAR 2013; 18 (9) : 85-98.
[3] Freund Y, Schapire RE. IJAR 2012; 10 (4) : 13-28.
[4] Hamid IRA, Abawajy JH. IJAR A 2010; 15 (2) : 10-28.