Detecting phishing website using Pattern Mining

Sreelekha B, Harika B, Mrs. L. Sujihelen
Department of Computer Science & Engineering,
School of Computing,
Sathyabama Institute of Science and Technology, Chennai
sujihelen@gmail.com

Abstract. Nowadays users are purchasing products and payments through online. While using these websites they will ask for the information of the users like username and password etc. There are some websites which are used to hack the details of the user. This type of websites is called phishing websites. For detecting these websites, they are proving an effective, intelligent data mining algorithm is used. In this paper, they are surveying the detection of phishing websites by using pattern mining technique.

Keywords— Novel algorithm, NLP technique, Supervised learning algorithm, streaming analytics, SVM, Moby fish.

1. Introduction

To hack and obtain the personal information from the user phishing technique is used. To overcome these phishing techniques there are many anti-phishing websites are used. Every day the usage of internet is increasing rapidly. Nowadays all the users are using internet for sending the messages, emails, and also online banking etc. From this, some hackers may ask for username and the password to collect the sensitive information of the users. Sometimes the users use the fake websites and give their information; they will target the main user of the particular bank or the financial organisation for collecting the information. For this purpose, they are providing the phishing technique to detect these types of websites and emails. In this without the use of internet connection we are going to detect the phishing websites. The pattern mining technique is used for detecting.

2. Related Work

“PHISHSTORM: DETECTING PHISHING WITH STREAMING ANALYTICS”

In this technique, we are going to use Phish Storm, which is a phishing detection used in real time for any URL to identify phishing sites. The experimental evidence will support this paper and is used to detect phishing websites. To evaluate this, we are going to use this concept which is Intra-URL related process which are extracted from words that compose the URL. Machine-learning-based classification features are used to detect Phishing URLs from real data set. In this paper, They are using efficient patterns which allow real-time analytics by using Big data diagrams such as STORM and advanced data structures are used.

A HYBRID MODEL TO DETECT PHISHING SITES USING SUPERVISED LEARNING ALGORITHM
Because of this, they can catch the credential details such as bank account number, password etc. To overcome this threat, They propose a hybrid algorithm. In this hybrid algorithm, select the best classifier in terms of high accuracy and less error rate. Then combine the best classifiers with others and develop the hybrid model.

In this, they use data from UCI repository. Datasets consist of attributes and instances including:
- Domain registration-length
- Using pop-up-window
- URL-of-anchor

Data splitting criteria
In this, the processed data is classified into training and testing categories. Some sets are used for training and others are used for testing.

Data mining classification technique
They use data mining classification techniques such as random forest, decision tree.

Ensemble technique
- Boosting
- Bagging
- Stacking

“Detecting Spam and Phishing Emails using SVM and Obfuscation URL Detection Algorithm”
They propose a hybrid model and evaluate their performance. In this, they proposed a system that uses SVM along with map reduce technique to overcome the problems related to SVM. They use BURL detection algorithm which provides multilayer security. Since/ the attackers use a number of ways to obfuscate the URL. OBURL detect the maximum number of attacks by using tests which are performing for checking
- NS test
- IP address test

SVM algorithm-support vector machine
Support vector machine can be applied to large amount of input data. It finds the separating hyperplane such that error for the unseen pattern is minimized.

A Novel Algorithm to Detect Phishing URLs
As they know that the use of the internet is increasing every day. Due to this many online attacks are increasing. The well-known attack is phishing. Phishing is the technique in which it is used to extract the delicate data of the users by providing the source in a web transaction. In many cases, phisher chooses HTTP URLs for the attack. For this purpose, they are proposing a Novel algorithm which will detect whether a given HTTP URL is a phishing site or not. The alert message will be displayed, if the URL found to be phishing, otherwise it will display the safe message. The algorithm enhances the performance when displaying with known/old phishing URLs.
In this, we are proposing phishing URL detection algorithm to find out maximum URL phishing attacks because it executes multiple tests such also URL features test. It also uses top Google a search engine results. The extent of our algorithm is limited to identify phishing URLs rather than detecting an extensive variety of malicious URL, rather than detecting an extensive variety of malicious URLs. In this algorithm, we are using URL hyperlink as input. The output will display the alert message whether the URL is phishing or not.

**DC SCANNER: DETECTING PHISHING ATTACK**

The proposed system uses the html data which consists of emails and also web pages. Also, domain and domain related authority to web pages are analyzed. There are two problems one is, a normal URL shown it in the address bar which appears like address of any website. Defining the work is meant for an email scanner to filter phishing by using the contents of web pages, also verifying the web domain and details.

Web page scanning

The deceive will come true when phishing web page having its normal appearance and URL same of the genuine web page, ask for credentials from users.

Script scanning

They can associate script codes. It is also checked whether the tags which follow the definitions are tortured.

**DETECTION AND PREVENTION OF POSSIBLE LOGIN ATTEMPTS THROUGH STOLEN CREDENTIALS FROM A PHISHING ATTACK IN AN ONLINE BANKING SYSTEM**

In this they detect and prevent unauthorized login attempts using behavioural based analysis, IP and device identification technologies.

Online banking facility

In order to access an online banking facility, usually a customer should first register with online banking service provide by the bank and setup username and a password.

a) Phishing attacks on online banking systems:

The concept of phishing is related to deceiving the customers to reveal their login credentials to attackers by masquerading as a trustworthy entity and the term first used by the spam people to describe about stealing of America Online (AOL) accounts.

b) Currently available methods to safeguard online banking:

The main action a financial institute is taking to protect their online customer is to educate the customer on online banking services security. The bank sometimes uses commercial products like Guardian Analytics, Threat Matrix, Actimize and the Ver safe Tot All to protect their customers from attackers.

c) Anomaly- based detection of frauds

Statistical methods-Detects the abnormalities of users or systems by monitoring the behaviour over a time period by measuring certain variables.

Distance-based-methods-More advanced than statistical methods and uses algorithms to detect outliers by computing distances among points.
"EFFECTIVE DEFENSE SCHEMES FOR PHISHING ATTACKS ON MOBILE COMPUTING PLATFORMS"

A lightweight anti-phishing scheme which is used for mobile platforms. Mobifish verifies the validity of applications, webpages and accounts by comparing with the actual identity. The performance of Moby fish is experimentally evaluated with phishing URLs and also legitimate URLs, also phishing applications. Hence the result is that Moby fish is very effective way for detecting Phishing attacks which are used on mobile phones.

A COMPUTER VISION TECHNOLOGY TO DETECT PHISHING ATTACKS"

The proposed approach is a combination of visual similarity-based techniques and whitelist. SURF detector which is the computer vision technique is used to extract features from both harmful and normal websites.

Some phishing has same visual appearance as the legitimate site such that we cannot differentiate with normal eyes. In proposed system we are working by detecting and extracting unique key points from both screenshots and it will finds the similarity ratio to identify phishing status.

This basic of the proposed solution is given below

1. Image database consisting of all URLs and popular websites screenshots has to be maintained.
2. After obtaining the accessed URL then compare with a whitelist of URLs.
3. URL is considered as innocent if the comparison is successful and no checking required.
4. In the whitelist if the given URL is not found then then apply SURF algorithm website screenshot and legitimate image database.
5. From both suspicious website screenshot, image database extract the SURF features and compare for similarity check.

1. Web page will be considered as suspected, if the similarity score is greater than the threshold
2. URL is considered to be innocent, if the similarity score is less than threshold. In next update the domain will be the part of whitelist.

3. A COMPARATIVE ANALYSIS AND AWARENESS SECURITY OF PHISHING DETECTION TOOLS

In this paper comparison of eight phishing detection tools has been done to find the best one by testing each tool on the data set, and further an awareness survey was carried out about these tools. After testing on the data set it is found that Antiphishing toolbar did a very good job at identifying 94.32% of phishing as well as legitimate websites from the data set.

Table 1. Study of existing work

|   | Detecting phishing attacks from URL by using NLP techniques (IEEE 2017) | Natural Language Processing (NLP) techniques. The system was implemented by examining URLs used in Phishing Attacks before opening them with using some extracted features | It is not predicted accurately |
|---|---|---|---|
| 1 | Detecting spam and phishing mails using SVM and obfuscation URL detection algorithm | a system that uses SVM technique along with map-reduce paradigm to achieve a higher accuracy in detection of the spam | Perhaps the biggest limitation of the support vector approach lies in the choice of the kernel |
3. Phish Storm: Detecting Phishing with Streaming Analytics
   (IEEE 2014)
   Phish Storm can interface with any email server or HTTP proxy, argue that phishing URLs usually have few relationships between the part of the URL that must be registered and the remaining part of the URL. They can be abysmally slow in test phase.

4. A Hybrid Model to Detect Phishing-Sites Using Supervised Learning Algorithms
   (IEEE 2016)
   To identify such features is a classification task and can be solved using data mining techniques. You can scam your emails or pop ups.

5. A novel algorithm to detect phishing URLs
   (IEEE 2016)
   A novel algorithm which will detect whether a given http URL is of phishing a site or not. Phishers can use the data to access a victim's account and withdraw money or purchase merchandise or services.

4. CONCLUSION

This paper focuses on various literature survey done to identify phishing websites. In this we are using pattern mining technique which is more effective when compared to the classification algorithm.

REFERENCES

[1] Ebubekir Buber, Banu Diri, Ozgur Koray Sahingoz, “Detecting phishing attacks from URL by using NLP techniques” IEEE 2017.

[2] Albert Mayan.J, Surya.B, Pranoy Prabhakar, Princekumar, ”Department–Student Library Using Twig Pattern Query Processing Over Admin-User Login Privilege”, Pak. J. Biotechnol. Vol. 13, Pp. 489 - 493, (2016).

[3] Prajakta Patil, Rashmi Rane, Madhuri Bhaulekar, “Detecting spam and phishing emails using SVM and obfuscation URL detection algorithm” 2017.

[4] Samuel Marchal, Jerome Francois, Radu State, Thomas Engel, “Phish Storm: Detecting Phishing with Streaming Analytics”, 4 dec 2014.

[5] Julian Menezes .R, Albert Mayan .J, M. Breezely George,” Development of a Functionality Testing Tool for Windows Phones”, Indian Journal of Science and Technology, Vol:8, Issue:22, pp: 1-7, September 2015, ISSN : 0974-6846.

[6] Varsharani Ramdas Hawanna, V.Y. Kulkami, RA.Rane, “A novel algorithm to detect phishing URLs”, 2016.

[7] Binay Kumar, Pankaj Kumar, Ankit Mundra, Shikha Kabra, “DC scanner: Detecting Phishing Attack”, IEEE 2015.

[8] Shammi Ishara Hewamadduma, “Detection and prevention of possible login attempts through stolen credentials”, IEEE 2017.

[9] Longfei, Xiaojiang DU,”Effective defense schemes on mobile computing platforms”, IEEE 2015.

[10] Routhu Srinivasa Rao, Syed Taqi Ali, “A computer vision technology to detect phishing attacks”, IEEE 2015.

[11] Himani Sharma, Er. Meenakshi, Dr. Sandeep Kaur Bhatia, “A comparative analysis and awareness survey of phishing detection tools”, IEEE 2017.
[12] Albert Mayan J, Kumar Akash Deep, Mukesh Kumar, Livingston Alvin Siva Prasad Reddy (2016), "Number plate recognition using template comparison for various fonts in MATLAB ", International Conference on Computational Intelligence and Computing Research, pp: 1-6.

[13] Sujihelen, L., Jayakumar, C., "Inclusive Elliptical Curve Cryptography (IECC) for Wireless Sensor Network Efficient Operations " Wireless Personal Communications, pp. 1-22, 2018.