INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

MALICIOUS URL DETECTION

[1.] Mr. T. LAYARAJA, [2.] S. SANTHOSH REDDY,[3.] CH. SHIVA KUMAR,

[4.] A. HARISH KUMAR

[1.] Assistant Professor Department of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Rangareddy (TS).India.

Email-:[1.] layaraaja@gmail.com

[2,3,4.]B.Tech StudentstDepartment of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Rangareddy (TS).India.

Email-:[2.]santhoshreddysoma4@gmail.com, [3.] cheni.shiva1509@gmail.com,

[4.]harishallakonda3@gmail.com.

Abstract- We address our experience training and testing a malicious URL detection system in this article. Our research is inspired by a range of technical and security developments. To begin with, the internet has become a more dangerous environment. Semanteme announced a 36 percent rise in cyber threats year over year in 2011. This equates to about 4,500 new attacks every day. The rate at which new attacks are launched has far outpaced the capabilities of conventional anti-malware tools. Second, both personal and business use of mobile web data has improved significantly. Semanteme observed in their 2012 State of Flexibility Survey that, while smartphones were once largely banned by IT, they are now used by hundreds and thousands of workers around the world. As a result, the attackable demographic for attackers has not only expanded, but also contains a potentially more appealing community from a commercial or financial perspective. We obtain a performance of 0.84 and a F1-measure of 0.74 using an Logistic Regression with a polynomial kernel. The user is, however, expected to take some action in all situations, such as click on a preferred resource on the internet (URL). The web security organizations have developed blacklisting programs to help identify malicious websites.

KEYWORDS: Malicious , URL, Attacks, Attackers, Internet.

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

1. INTRODUCTION

New connectivity tools have had a huge effect on company development and promotion through a wide range of applications, including online banking, u t, and instant messaging. In reality, having an internet presence is almost required to operate a successful business in today's world. As a result, the Massive Global Web's value has been steadily growing. Regrettably, technical advances are accompanied by new advanced tactics for attacking and defrauding people. Rogue web-sites that has capacity to sell the counterfeit a goods, the financial manipulation that fool users in to exposing confidential details that leads to themoney or the identification of theft, and even an malware installation in the most users device is examples of such attacks.

Usually, phishing attacks use sql injection to deceive the user into clicking on a spoofed connection that leads to a false web page. The spoof connection is posted on famous websites or sent to the victim by email. The false website is designed to look like the real one. As a result, rather than sending the victim's request to the actual web application, it will be sent to the fake web server.

PROBLEM DESCRIPTION

Let's look at the URL layout to get a clearer sense of what attackers are considering asthey create a phishing domain. To address web sites, the Universal Resource Locator (URL) was developed. The diagram below illustrates the related parts of a standard URL's structure. It all starts with the protocol that is used to reach the website.

Security threats of malicious URL's

Malicious websites are a widespread and severe cybersecurity problem. Malicious URLs hosts the unsolicited and irrelevant content (malware, spoofing, driven by hacks, and so on) and trick unwitting consumers into becoming scam victims (monetary loss and theft of the privatized information and the malware installation), resulting in billion

ISSN: 2057-5688

dollars in losses per years. It was critical for identifying and respond to certain threatens as soon as possible.

Scope

The reach of this strategy is restricted to the host. If the host is untrustworthy but the URL is secure, it will also be labelled as malicious due to the host, resulting in a false positive. However, if the URL is malicious and hosted by a well-known host, it could be misclassified as benign, resulting in a false negative.

2. LITERATURE SURVEY

CANTINA

Centered on the TF IDF informative retrieve algorithm, Hong et al proposed a content- based methods for finding phish websites. The design and methodology of a few heuristics are also enlightened in this paper. It was created in-order to reduces the count of false cases. The results of this study enhance CANTINA is capable of identifying phishing sites, effectively labelling about 95% of phishing targets.

CANTINA, 1 the novel based on technique for detecting phishing-sites, was implemented at with the preparation, execution, and evaluation. Unlike other methodologies that looks a surface attributes of a site page, such as the URL, domain name, CANTINA looks at the idea of a site to determine if it is genuine or fake. CANTINA make open of the known TF-IDF formula actually, the Robust calculation recently developed by Phelps and Wilensky.

CANTINA+

Cranor et al. proposed CANTINA+, an element rich AI scheme that aims to use AI to exploit the expressiveness of a rich array of highlights in order to gain the high Accurate Positive rate (AP), on novels phishes while limits the False Positive rate to the lowest level using sifting calculations. CANTINA+, most used element bases approach in the writing, which elaborates the HTML Documentation Objective Model (DOM), web engine tools, outsider administrations using AI technique for identification of phishes, includes eight novel highlights. They devised two channels to aid in the reduction of FP. The firstly are a close-copy phish locator that hashing to generate phish that is extremely similar. The second is a login framework channel, which categorizes Web pages that have no known

ISSN: 2057-5688

login structure as genuine. At last, CANTINA+ has been demonstrated for being a serious phishing adversary.

3. EXISTING SYSTEM:

Many studies for detecting and identifying malicious URLs have been proposed in recent years to detect and prevent malicious URL attacks.

In general, such URLs are blacklisted in utility software, however, in the event of a new URL, utility software is unable to identify if it is harmful.

As a result, there is a need for a system that can detect a new malicious URL. Machine learning techniques can be used to protect against these threats.

DISADVANTAGES OF EXISTING SYSTEM:

It cannot detect new urls other from the blacklisted urls.

Manual detection of malicious urls are needed.

4. PROPOSED SYSTEM:

In this project, ML is employed using the WEKA tool. This tool provides a set of visualization techniques and methodologies for data analysis and predictive modeling.

It has numerous ML classifiers which can be used to detect malicious URLs.

The ISCX-URL-2016 dataset from kaggle.com is used for evaluation purposes.

The ML classifiers considered in this project is logistic regression.

ADVANTAGES OF PROPOSED SYSTEM:

Because of machine learning new urls are effectively detected.

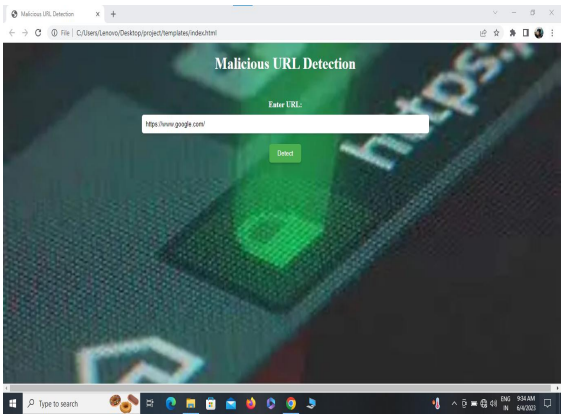
Because of effective dataset and logistic regression gives 84% accuracy.

5. MODULES:

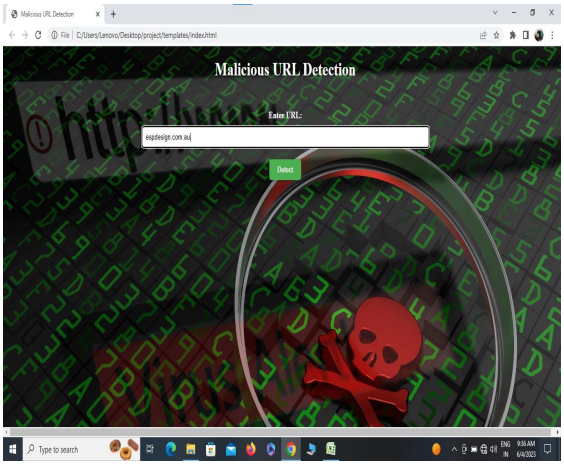
1. Pandas
2. Sklearn(model_selection, linear_model, metrics, impute)
3. Numpy
4. Urllib.parse
5. Tldextract

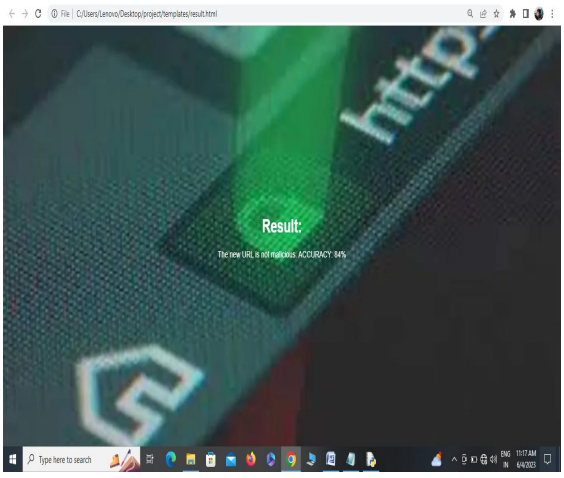
6. RESULTS:

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688



Input 1





Output1

Output2

7. CONCLUSION

Many cyber security applications depend on malicious URL identification, and machine learning techniques are obviously a promising path. We conducted a thorough and ordered analysis of Malicious Detection using AI approaches in the work. We provided the methodical description of Malicious detection from an AI standpoint, followed by nitty gritty information. Current investigations finds malicious URL identification, especially in the types of growing new component portrayals and preparing new learning calculations for determining vindictive URL position assignments We sorted most, if not all, of the existing obligations for malevolent URL position in writing in this overview, as well

ISSN: 2057-5688

as acknowledged the requirements and challenges for creating tasks for detecting malicious URLs. In this analysis, we summarize the majority, if not all, of the existing commitments for malignant URL location in writing, as well as the requirements challenges for the develop of Malicious Detection as the Services for the real-world cyber security applications.

8. REFERENCES

1. Abdelhamid N, Ayesh A, Thabtah F (2014) Phishing detection based associative classification data mining. Science-Direct 41:5948–5959.
2. Chen KT, Chen JY, Huang CR, Chen JY (2009) Fighting phishing with discriminative key point features of webpages. IEEE Internet Comput 13:56–63.
3. Chen X, Bose I, Leung ACM, Guo C (2011) Assessing the severity of phishing attacks: a hybrid data mining approach. Expert Syst Appl 50:662–672.
4. Fu AY, Wenyin L, Deng X (2006) Detecting phishing web pages with visual similarity assessment based on earth mover's distance. IEEE Trans Dependable Secure Comput 3(4):301 321.
5. Islam R, Abawajy J (2013) A multi-tier phishing detection and filtering approach. J NetwComput Appl 36:324–335.
6. Li Y, Xiao R, Feng J, Zhao L (2013) A semi-supervised learning approach for detection of phishing webpages. Optik 124:6027– 6033.
7. Nishanth KJ, Ravi V, Ankaiah N, Bose I (2012) Soft computing-based imputation andhybrid data and text mining: the case of predicting the severity of phishing alerts. Expert Syst Appl 39:10583–10589.