Chapter 19
Prediction and Prevention of Malicious URL Using ML and LR Techniques for Network Security: Machine Learning

S. Mythreya
Koneru Lakshmaiah Education Foundation, India

A. Sampath Dakshina Murthy
https://orcid.org/0000-0002-9960-6373
Vignan's Institute of Information Technology, India

K. Saikumar
https://orcid.org/0000-0001-9836-3683
Koneru Lakshmaiah Education Foundation, India

V. Rajesh
Koneru Lakshmaiah Education Foundation, India

ABSTRACT

Understandable URLs are utilized to recognize billions of websites hosted over the present-day internet. Opposition who tries to get illegal admittance to the classified data may use malicious URLs and present them as URLs to users. Such URLs that act as an entry for the unrequested actions are known as malicious URLs. These wicked URLs can cause unethical behavior like theft of confidential and classified data. By using machine learning algorithm SVM, we can detect the malicious URLs. One of the essential features is to permit the benevolent URLs that are demanded by the customer and avoid the malicious URLs. Blacklisting is one of the basic and trivial mechanisms in detecting malicious URLs.

DOI: 10.4018/978-1-7998-9640-1.ch019

INTRODUCTION

A Uniform Resource Locator (URL) naturally termed as web address, reference to a web resource that specify its position on a computer network. URL has exact composition and arrangement. Assailant often alters one or added elements of the structure of the URL to betray used for scattering their malicious URLs. Malicious URLs are the links influence the customers by clicking on an infected URL, users can download ransom are, virus or any type of malware that will compromise user's machine or even their network (Odeh et al., 2021). A malicious URL is just a link that, when clicked, takes the user to a harmful website or page on the internet. As the name says, a malicious URL can do nothing but harm. To accomplish malicious objectives like as advancing a political agenda, stealing confidential information about individuals or businesses, or just making a fast money is the usual motivation for rogue web pages. Fake and authentic websites may both include dangerous links, which should be taken into consideration. A cybercriminal scan may either produce a completely phony and harmful website, or it can create dangerous URLs for legal websites. Drive-by-downloads, phishing, and other forms of social engineering and spam are just a few of the ways malicious URLs are spread. According to statistics, attackers using spreading malicious URLs ranked first among 10 attack techniques (Wanda & Jie 2019). The three major URL scattering methods are malicious URLs, both net URLs and phishing. From the enlargement in the number of malicious URL distributions over the successive days, it is apparent that there is necessitate learning and applying methods to identify and avoid the malicious URLs. To identify and avoid malicious URLs we used machine learning algorithm Logistic Regression. By using this algorithm, it is Easy to detect malicious URL and get rid of malware activities.

Phishing is an online social designing assault that intends to take an individual's computerized personality by imitating a trustworthy substance. The aggressor sends an assault vector, which can be an email, a visit meeting, a blog entry, or whatever else, that contains a connection (URL) to a malevolent site that is utilized to inspire individual data from the people in question. We are especially keen on fostering a framework for URL examination and grouping to forestall phishing attacks. Rather than getting to the site and acquiring highlights from it, URL investigation is interesting to keep the distance between the aggressor and the person in question. It's additionally quicker than the Internet look as far as recovering substance from the objective site and organization level properties, which were used in earlier examinations (Fang et al., 2021). We investigate an assortment of parts of URL examination, remembering execution examination for both adjusted and uneven datasets in both a static and live exploratory setting, just as online versus clump learning. Because of the consistently changing nature of attacks and the design of the present Web pages, detecting malicious Web pages has become a critical responsibility. Attackers use a variety of attack construction tactics. As a result, feature selection and dataset preparation are crucial for detecting fraudulent Web sites. While existing technologies offer a potential answer for detecting fraudulent Web pages, there are still gaps in the detection process. We have introduced a static examination of URL strings for successful recognizable proof of pernicious Web pages in this work. Just the static parts of Web page URLs have been thought of. From benign and malicious URL benchmarks, we collected 79 static properties of URLs and domain names. On our dataset, we tested Support Vector Machine (SVM), AdaBoost, J48, Random Forest (RF), Random Tree (RT), Naive Bayes (NB), Logistic Regression (LR), SGD, & BayesNet batch learning algorithms. For all of the classification models, our experimental research indicates encouraging detection results, with a detection rate of 95 percent to 99 percent and a very low false positive rate (FPR) and false negative rate (FNR).