International Journal of Future Generation Communication and Networking
Vol. 13, No. 3s, (2020), pp. 782–784

Malicious Urls Detection Using Convolutional Neural Networks and Deep Learning (Cgru)

Shweta Badagu
IT Department
SITS, Narhe
SPPU, Pune
shwetabadagu45@gmail.com

Shraddha Gahandule
IT Department
SITS, Narhe
SPPU, Pune
shraddha98g@gmail.com

Apurva Jatala
IT Department
SITS, Narhe
SPPU, Pune
apurvajatla1997@gmail.com

Abstract

The World Wide Web supports a good range of criminal activities like spam advertised e-commerce, financial fraud and malware dissemination. Nowadays Machine learning and Deep learning algorithms are a popular trend for the security aspects. Activity which requires the user to take some action, like clicking on a desired Uniform Resource Locator (URL) may result in some output This project aims to create a Web-App for the detection of malicious URLs based on characters as text classification features.Since the malicious keywords are unique to URLs, use of feature representation method of URLs based on malicious keywords will be used.

Keywords: -Machine learning, CNN, GRU, User Privacy, Intent Publishing, Security, URL.

I. INTRODUCTION

With the dominance of the World Wide Web as the platform supporting knowledge and increased economic activity, the security aspect continues to be at the forefront of many companies and governments' research eff orts. In order to spot these malicious sites, the online security community has developed blacklisting services. These blacklists are in turn developed by a mix of techniques including manual reporting, honeypots, and web crawlers combined with site analysis. While URL blacklisting has been eff ective to some extent, it's rather easy for an attacker to deceive the system by slightly modifying one or more components of the URL string. Inevitably, many malicious sites aren't blacklisted either because they're too recent or were never or incorrectly evaluated. A large number of web applications can construct executable commands, SQL injections, XSS and other web attacks simply by embedding executable code or malicious code in URLs, so detection of malicious URLs is important for intrusion detection. In this model we are using the 2D tensor for encoding of URL, these encoded tensors will pass through the convolutional neural network, which will derive a sigmoid. This Sigmoid will then pass through a Gated Recurrent Unit which will accept a reset and update gate to predict the malicious content.

II. RELATED WORK

With the continual development of Web attacks, many web applications are affected by various sorts of security threats and network attacks.The security detection of URLs has always been the focus of Web security.An attacker can construct various web attacks such as SQL, XSS, and information disclosure by embedding executable code or injecting malicious code into the URL.

It's very important to be reliable on various security web applications to detect malicious URLs. This project designs a convolutional gated recurrent neural network for keyword based text classification.

ISSN: 2233-7857 IJFGCN

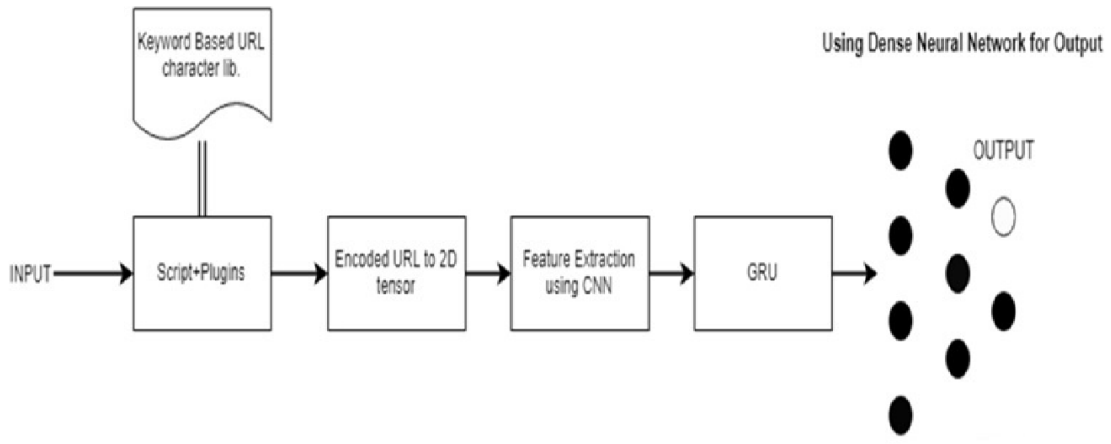
Some of the commonly used work to detect malicious URLs in the field of network security are mainly described by methods such as blacklist-based detection methods, traditional machine learning methods and deep learning methods based on feature extraction. After reviewing existing techniques, we propose a detection model that combines a convolutional neural network and gated recurrent unit (GRU).
We are going to implement a convolutional neural network with a gated recurrent unit and a certain python based API for analyzing the URL Dataset.
And we are using the pre filtering techniques for analysing URLs.
The use Logistic Regression where it can take only two values, "0" and "1" which we implement in a gated recurrent unit.

III. MOTIVATION

The World Wide Web supports a good range of criminal activities like spam-advertised e-commerce, financial fraud and malware dissemination. Although the precise motivations behind these schemes may differ, the common denominator lies within the incontrovertible fact that unsuspecting users visit their sites. These visits are often driven by email, web search results or links from other sites . In all cases, however, the user is required to require some action, such as clicking on a desired Uniform Resource Locator (URL). These URLs about its structure the naïve user click should be malicious free.

IV. SYSTEM ARCHITECTURE

This System is the overview of the working CGRU model for malicious URL detection. This model combines the characteristics of URLs in the field of Web attacks at the character level. It uses convolutional neural networks and gated recurrent units to extract features from URLs. Finally, it combines the classification module to detect Web attacks.
Figure 1 The Proposed System



ISSN: 2233-7857 IJFGCN

V. CONCLUSION

This system introduces a neural network model CGRU for malicious URL detection in the field of cyber security. Using experiment results of previous existing techniques for manually extracting features for malicious URLs and comparing experiments of other classification models, we proposed our system, using our model it has a good effect of malicious URL detection.

REFERENCE

[1] P. Prakash, M. Kumar, R. R. Kompella, and M. Gupta, "Phishnet: Predictive blacklisting to detect phishing attacks," in 2010 Proceedings IEEE INFOCOM. San Diego, CA, USA: Citeseer, 14-19 Mar 2010, pp. 1–5.

[2] D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL detection using machine learning: A survey," arXiv:1701.07179 [cs.LG], 2017.

[3] J. Saxe and K. Berlin, "eXpose: a character-level convolutional neural network with embeddings for detecting malicious URLs, file paths and registry keys," arXiv:1702.08568 [cs:CR], 2017.

[4] B. Cui, S. He, X. Yao, and P. Shi, "Malicious url detection with feature extraction based on machine learning," International Journal of High-Performance Computing and Networking, vol. 12, no. 2, 2018, DOI:10.1504/IJHPCN.2018.10015545.

ISSN: 2233-7857 IJFGCN