VISHNU
UNIVERSAL LEARNING

BVRIT HYDERABAD College of Engineering for Women

Department of Information Technology

Detection of Malicious Data Through Attacks

Under the Guidance of:

Guide Name: Dr.SL.Aruna Rao

Designation: Professor & HOD

Team- 3

Anuhya Javvaji (18WH1A1240)

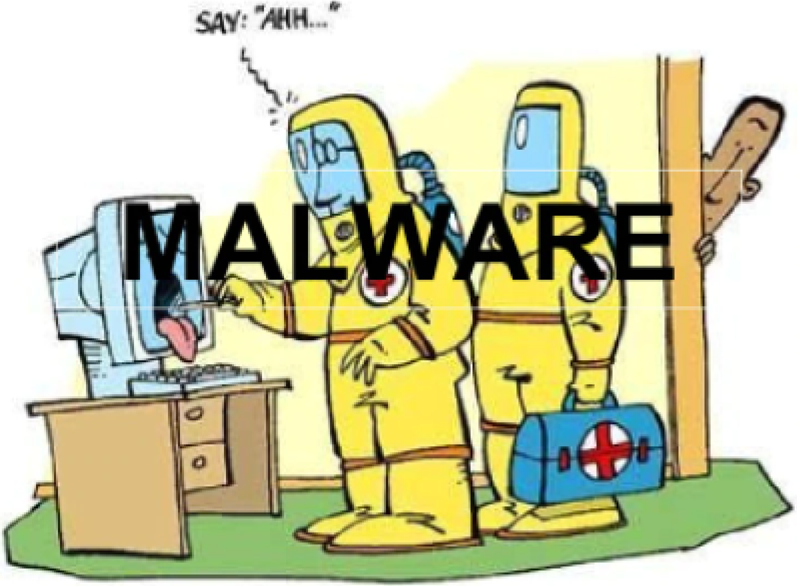
A.Meher Gayatri Devi (18WH1A1206)

T. G.S.N Sai Chandana(18WH1A1207)

Contents

- Abstract
- Introduction
- Problem Statement
- Literature Survey
- References

MALWARE



Abstract

While the World Wide Web has become a killer application on the Internet, it has also brought in an immense risk of cyber-attacks. The more we grew in technology over internet the more we get exposed to malicious data which led to cyber-attacks. With the rise in the underground Internet economy, automated malicious programs popularly known as malwares have become a major threat to computers and information systems connected to the internet. Malicious URLs have been widely used to mount various cyber-attacks including spamming, phishing and malware. We propose a machine learning model that takes a dataset, train and then distinguish good data and bad data.

Introduction

- The security issue arises along with the evolution of network, the diversity of network services and applications provides hackers more opportunities to compromise the network than ever before.
- To protect networks against malicious access is always challenging even though it has been studied for a long time. Due to the evolution of network in both new technologies and fast growth of connected devices, network attacks are getting versatile as well.
- Malicious URL is a link created with the purpose of promoting scams, attacks, and frauds. By clicking on an infected URL, you can download ransomware, virus, trojan, or any other type of malware that will compromise your machine or even your network, in the case of a company.

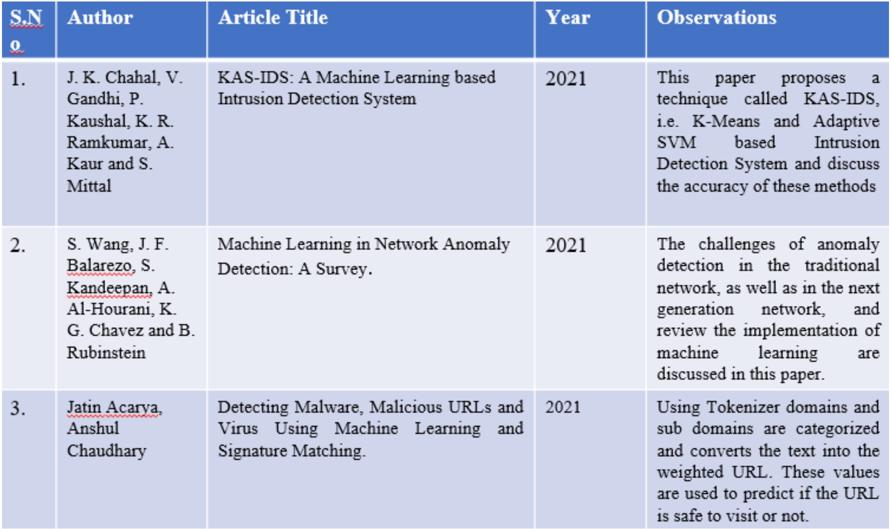
Problem Statement

Malware is any software intentionally designed to cause damage to a computer, server, client or network by tampering the data in the computer network. By contrast, software that causes unintentional harm due to some deficiency is typically described as a software bug. A malicious URL is short and simple but can cause a lot of damage and the potential harm is so big that malicious links are considered one of the biggest threats to the digital world.

Problem Definition

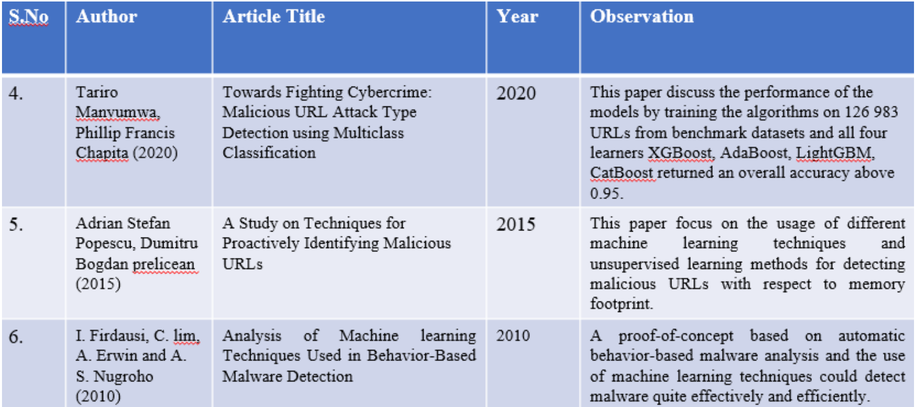
We use different machine learning algorithms that takes a dataset, train and then distinguish good data and bad data. The core of ML models' detection efficiency relies on the dataset's quality to train the model. It can also be used to detect malware variants.

Literature Survey



S.No	Author	Article Title	Year	Observations
1.	J. K. Chahal, V. Gandhi, P. Kaushal, K. R. Ramkumar, A. Kaur and S. Mittal	KAS-IDS: A Machine Learning based Intrusion Detection System	2021	This paper proposes a technique called KAS-IDS, i.e. K-Means and Adaptive SVM based Intrusion Detection System and discuss the accuracy of these methods
2.	S. Wang, J. F. Balarezo, S. Kandeepan, A. Al-Hourani, K. G. Chavez and B. Rubinstein	Machine Learning in Network Anomaly Detection: A Survey.	2021	The challenges of anomaly detection in the traditional network, as well as in the next generation network, and review the implementation of machine learning are discussed in this paper.
3.	Jatin Acarya, Anshul Chaudhary	Detecting Malware, Malicious URLs and Virus Using Machine Learning and Signature Matching.	2021	Using Tokenizer domains and sub domains are categorized and converts the text into the weighted URL. These values are used to predict if the URL is safe to visit or not.

Literature Survey



S.No	Author	Article Title	Year	Observation
4.	Tariro Manyumwa, Phillip Francis Chapita (2020)	Towards Fighting Cybercrime: Malicious URL Attack Type Detection using Multiclass Classification	2020	This paper discuss the performance of the models by training the algorithms on 126 983 URLs from benchmark datasets and all four learners XGBoost, AdaBoost, LightGBM, CatBoost returned an overall accuracy above 0.95.
5.	Adrian Stefan Popescu, Dumitru Bogdan prelicean (2015)	A Study on Techniques for Proactively Identifying Malicious URLs	2015	This paper focus on the usage of different machine learning techniques and unsupervised learning methods for detecting malicious URLs with respect to memory footprint.
6.	I. Firdausi, C. lim, A. Erwin and A. S. Nugroho (2010)	Analysis of Machine learning Techniques Used in Behavior-Based Malware Detection	2010	A proof-of-concept based on automatic behavior-based malware analysis and the use of machine learning techniques could detect malware quite effectively and efficiently.

References

- https://ieeexplore.ieee.org/document/5675808
- https://ieeexplore.ieee.org/document/9609402
- https://ieeexplore.ieee.org/document/9610045
- https://arxiv.org/abs/2105.13435
- https://ieeexplore.ieee.org/document/9456440
- https://ieeexplore.ieee.org/document/9396014
- https://medium.com/cuelogic-technologies/ evaluation-of-machine-learning-algorithms-for-intrusion-d

Thank
you