"JPEGVigilant: AI-Powered Malware Image Detection"

A SYNOPSIS SUBMITTED

to

SAVITRIBAI PHULE PUNE UNIVERSITY

FOR THE PARTIAL FULFILMENT OF AWARD OF DEGREE

BACHELOR OF ENGINEERING

in

COMPUTER ENGINEERING

SUBMITTED BY

1. Mr.Kharade Nilesh Shahaji	**Roll No: 69**
2. Mr.Korake Digvijay Dilip	**Roll No: 73**
3. Mr.Kshirsagar Dipak Vinod	**Roll No: 74**
4. Mr.Mind Rushikesh Satish	**Roll No: 88**

UNDER THE GUIDANCE OF

Prof.Ekatpure J.N.

DEPARTMENT OF COMPUTER ENGINEERING

SPVP's SBPCOE, Indapur

DEPARTMENT OF COMPUTER ENGINEERING

SPVP'S S. B. PATIL COLLEGE OF ENGINEERING

VANGALI, INDAPUR, PUNE 413106

2023 - 24

SYNOPSIS

---

1. Project Group Information

Group Id:12

Name of Students:

1. Mr.Kharade Nilesh Shahaji
2. Mr.Korake Digvijay Dilip
3. Mr.Kshirsagar Dipak Vinod
4. Mr.Mind Rushikesh Satish

2. Project Title:

"JPEGVigilant: AI-Powered Malware Image Detection"

3. Project Sponsorship details

Sponsorship Company:NA

External Guide: NA

Sponsership Company Address: NA

4. Problem Statement

To build and implement web application for Machine Learning Based Solution for the Detection of Malicious JPEG Images

5. Area of Project

Image Processing With AI

6. Abstract

Cyberattacks against people, companies, and organisations have risen in recent years.In order to conduct an attack, cybercriminals are constantly searching for efficient channels to spread malware to targets. Millions of people use photos every day, and the majority of consumers believe that they are safe to use. However, some types of images may contain malicious payloads that carry out dangerous functions. Due in large part to its lossy compression, JPEG is the most widely used image format.In this study, we introduce MalJPEG, the first machine learning-based method designed exclusively for the quick and accurate identification of unknown malicious JPEG images. In order to distinguish between benign and malicious JPEG images, MalJPEG statically derives 10 straightforward yet discriminative properties from the JPEG LE structure.

7. Goals and Objectives

Goals: The major goal of this system focuses on the use of machine learning for the identification of fraudulent images, especially JPEG images.

Objectives:

- To train a machine learning model, you would need a diverse dataset of both benign (non-malicious) and malicious JPEG images.
- To study Voice-Enabled Traffic Sign Recognition and Alert System using ML.
- To scan incoming JPEG images for malicious content.
- To refine the results and reduce false positives.

8. Relevant mathematics associated with the Project :

What do we need to find?

We need to find if there is any malicious JPEG image in the system. Let us consider S as a system for Courier Management System.

Input:

- Files containing malicious JPEG images.

Output:

- The software should detect the malicious JPEG image if the input file contains any

Set of functions:

a) Uploading the files of JPEG images

b) Scanning the file for presence of any Malicious JPEG images

c) Detecting Malicious JPEG images

Space Complexity:

The space complexity depends on Presentation and visualization of discovered patterns. More the storage of data more is the space complexity.

Time Complexity:

Check No. of patterns available in the datasets= n

If (n(1)) then retrieving of information can be time consuming. So the time complexity of this algorithm is O (nˆn).

Failures and Success conditions. Failures:

1. Huge database can lead to more time consumption to get the information.
2. Hardware failure.
3. Software failure.

Success:

1. Search the required information from available in Datasets.
2. User gets result very fast according to their needs.

9. Names of Conferences / Journals where papers can be published

1.IEEE

2.UGC Care

3.Springer

10. Review/ Literature survey of Conference/Journal Papers supporting Project idea

Sr. No	**Paper Title**	**Author**	**Year**	**Problem solved in this paper : Existing Problem Statement**	**Technique used to solve problem : Existing Problem Solution**	**What will be future work : Future Scope**
1	A Novel Machine Learning Approach for Malware Detection	Tarun Kumar, Sanjeev Sharma, Himanshu Goel, Sumit Cahudhary,Parag Jain	2019	This study A Novel Machine Learning Approach for Malware Detection.	we have proposed a framework for malware analysis based on semi automated malware detection usually machine learning which is based on dynamic malware detection.	Improve accurancy.
2	Detection of Advanced Malware by Machine Learning Techniques	Sanjay Sharma, C. Rama Krishna and Sanjay K. Sahay	2019	This study Detection of Advanced Malware by Machine Learning Techniques.	we study the frequency of opcode occurrence to detect unknown malware by using machine learning technique.	In future, we will implement proposed approach on different datasets and will perform in the deep analysis for the classification of advanced malicious software.

3	Novel active learning methods for enhanced PC malware detection in windows OS	Nir Nissim, Robert Moskovitch, Lior Rokach, Yuval Elovici	2019	We study Novel active learning methods for enhanced PC malware detection in windows OS	In this paper we proposed a framework based on new active 1274 learning methods (Exploitation and Combination) designed for 1275 acquiring unknown malware.	In future work, we are interested in implementing this frame- 1361 work also on Android applications where it is not very feasible to 1362 apply advanced detection techniques over the device itself due to 1363 its resource limitations (CPU, battery, etc.).

4	TrustSign: Trusted Malware Signature Generation in Private Clouds Using Deep Feature Transfer Learning.	Daniel Nahmias, Aviad Cohen, Nir Nissim, Yuval Elovicia	2019	To obtain and develope TrustSign: Trusted Malware Signature Generation in Private Clouds Using Deep Feature Transfer Learning. .	This paper presents TrustSign, a novel, trusted automatic malware signature generation method based on high-level deep features transferred from a VGG-19 neural network model pre-trained on the ImageNet dataset.	First direction for future work is related to maintaining the updatability and efficiency of our proposed solution.

5	Keeping pace with the creation of new malicious PDF files using an active-learning based detection framework.	Nir Nissim, Aviad Cohen, Robert Moskovitch, Asaf Shabtai, Matan Edri, Oren BarAd and Yuval	2019	To develope Keeping pace with the creation of new malicious PDF files using an active-learning based detection framework.	In this study we present an active learning (AL) based framework, specifically designed to efficiently assist anti-virus vendors focus their analytical efforts aimed at acquiring novel malicious content.	In future work, in addition to additional types of malicious documents we are interested in extending this framework to Android applications.
6	A Novel Machine Learning Approach for Malware Detection	Tarun Kumara, Sanjeev Sharmaa, Himanshu Goela, Sumit Chaudharya	2019	To obtain and develope A Novel Machine Learning Approach for Malware Detection	In this paper, we have proposed a framework for malware analysis based on semi automated malware detection usually machine learning which is based on dynamic malware detection.	Improving accuracy.

7	Dynamic Malware Analysis in the Modern Era—A State of the Art Survey	Nir nissim, aviad cohen1, jian wu, andrea lanzi, lior rokach, Yuval elovici and lee giles	2019	To study Survey of Dynamic Malware Analysis in the Modern Era—A State of the Art Survey.	In this study, we present related vulnerabilities and malware distribution approaches that exploit the vulnerabilities of scholarly digital libraries.	In future work, we suggest evaluating the malicious PDF presence in additional digital libraries such as MAS, Web of Science, and PubMed, as well as investigating them for vulnerabilities.
8	Dynamic Malware Analysis in the Modern Era—A State of the Art Survey.	ORI OREMEIR, NIR NISSIM, YUVAL ELOVICI, and LIOR ROKACH	2019	To study Survey of Dynamic Malware Analysis in the Modern Era—A State of the Art Survey.	We describe the advancements made in analysis techniques during this time. Early research centered on function call analysis, execution control, and flow tracking.	future research stems from the fact that dynamic analysis produces a time sequence output of observed behavior.

9	Survey of Machine Learning Techniques for Malware Analysis	Daniele Ucci, Leonardo Aniello, Roberto Baldoni	2018	To study Survey of Machine Learning Techniques for Malware Analysis.	This survey aims at providing an overview on the way machine learning has been used so far in the context of malware analysis in Windows environments.	The novel concept of malware analysis economics can encourage further research directions, where appropriate tuning strategies can be provided to balance competing metrics (e.g. accuracy and cost) when designing a malware analysis environment.

10	Malware Detection on Byte Streams of PDF Files Using Convolutional Neural Networks.	Young-Seob Jeong , Jiyoung Woo , and Ah Reum Kang	2018	To study Survey of Malware Detection on Byte Streams of PDF Files Using Convolutional Neural Networks.	In this paper, we design a convolutional neural network to tackle the malware detection on the PDF files.We collect malicious and benign PDF files and manually label the byte sequences within the files.	As a future work, we will collect data of other file types (e.g., .rtf files) and perform further investigation.

11. Plan of Project Execution

- Start date From Group Registration 6 July 2023 TO 31 March 2024.
- Submission of Synopsis 31 Aug. 2023.
- Submission of Survey Paper 11 September 2023.
- Submission of Design Paper 30 September 2023.
- Last date to submit Report SEM I Spiral 20 Oct. 2023.
  of Result Paper Feb. 2024.
- Last date to submit Black Golden Embossing Report SEM II – 15 March 2024.
- Completion of all documents and records till 15 March 2024 is must.
- All activities must be completed on or before above dates.

Guide	Project Coordinator	HOD
Prof.J.N.Ekatpure	**Dr.A.B.Gavali**	**Dr.S.T.Shirkande**