IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Ransomware Detection Using Machine Learning

Ms Suma K [1],Shashank R M[2], Skanda B N[3], Srujan S[4], Vijeth Shetty[5]

Assisstant Professor, Dept. of Computer Science & Engineering, Mangalore Institute of Technology & Engineering, Moodabidri, India[1]

Student, Dept. of Computer Science & Engineering, Mangalore Institute of Technology & Engineering, Moodabidri, India[2,3,4,5]

Abstract: Ransomware attacks pose significant security threats to personal and corporate data and information. The owners of computer-based resources suffer from verification and privacy violations, monetary losses, and reputational damage due to successful ransomware assaults. As a result, it is critical to accurately and swiftly identify ransomware. Numerous methods have been proposed for identifying ransomware, each with its own advantages and disadvantages. The main objective of this research is to discuss current trends in and potential future debates on automated ransomware detection. This document includes an overview of ransomware, a timeline of assaults, and details on their background. It also provides comprehensive research on existing methods for identifying, avoiding, minimizing, and recovering from ransomware attacks. An analysis of studies between 2017 and 2022 is another advantage of this research. This provides readers with up-to-date knowledge of the most recent developments in ransomware detection and highlights advancements in methods for combating ransomware attacks. In conclusion, this research highlights unanswered concerns and potential research challenges in ransomware detection.

Index Terms - machine learning, ransomware techniques, cybersecurity, ransomware detection, ransomware attacks.

I. INTRODUCTION

The rapid proliferation of ransomware attacks has emerged as one of the most significant cybersecurity threats facing organizations today. In recent years, ransomware has become an increasingly popular tool with which cybercriminals extort money from victims by encrypting their data and demanding payment for a decryption key. The impact of ransomware attacks has been felt across all industries, from healthcare and finance to government and education. Given the high stakes involved, it is crucial to understand the nature of ransomware attacks, how they spread, and the potential consequences of falling victim to one [1]. The importance of research in this area cannot be overstated. With the threat of ransomware attacks continuing to grow, there is a pressing need for scholars and practitioners to delve deeper into the problem and identify effective strategies for prevention and mitigation. This paper aims to contribute to this effort by providing a comprehensive overview of the ransomware threat landscape, analyzing the factors that contribute to the spread of ransomware, and exploring potential avenues for future research. By shedding light on this critical issue, we hope to help individuals and organizations better-protect themselves against ransomware attacks and mitigate the potential damage caused by these malicious programs [1].

II. LITERATURE SURVEY

[1] This study explains the potential use of various machine learning algorithms like Decision tree, Random forest and Neural networks. It also shows the idea of the accuracy ranges with respect to the following datasets.
[2] This study concentrated on the evolution of the ransomware attacks in various perspectives and the identifies the gaps in existing datasets and the feature engineering techniques which results in the use of hybrid model for the detection mechanism.
[3] The study shows the cognitive science principles applied to machine learning for realtime ransomeware detection and the classification.
[4] This paper will discuss the current trends of detection of the ransome file and the use of various algorithms like SVM, Decision tree using AI which helps in the easy and accurate detection of the breach of the user data.

III. SCOPE AND METHODOLOGY

3.1 Scope

Key Features of machine learning in ransomware detection include: File Behavior: Monitoring file-related activities such as file creation, modification, and encryption can help detect ransomware. Features related to the rate of file changes and unusual file extensions are important. The scope of ransomware detection using machine learning encompasses several critical areas, reflecting the increasing complexity and prevalence of ransomware attacks in today's digital landscape. As ransomware continues to evolve, traditional detection methods have proven inadequate, necessitating the adoption of advanced techniques. Machine learning (ML) offers a promising approach by enabling automated detection and classification of ransomware based on its dynamic characteristics. This capability is essential for developing robust cybersecurity measures that can adapt to new variants and attack strategies.

Research in this field focuses on various machine learning algorithms, including Decision Trees, Random Forests, Naïve Bayes, and Neural Networks, to enhance detection accuracy and efficiency. The integration of feature selection techniques further improves model performance by identifying the most relevant attributes for classification. Additionally, there is a growing emphasis on creating comprehensive datasets that accurately represent both benign and malicious behaviors, which is crucial for training effective models.
The scope also includes exploring hybrid approaches that combine multiple machine learning techniques to leverage their strengths. As the threat landscape evolves, continuous research is needed to address vulnerabilities in existing models and develop innovative solutions that can preemptively detect ransomware before it causes significant damage. Overall, the field of ransomware detection using machine learning holds substantial potential for improving cybersecurity defenses and protecting sensitive data across various sectors.

3.2 Methodology

Ransomware has become a critical threat in cybersecurity, targeting individuals and organizations by encrypting data and demanding ransom for its release. Traditional detection methods often fall short against sophisticated ransomware variants, prompting the need for advanced solutions. Machine learning (ML) has emerged as a powerful tool for automating the detection of ransomware by analyzing dynamic characteristics of malware. This approach allows for the classification of malicious software with greater accuracy and efficiency compared to conventional methods.

The methodology for ransomware detection using machine learning involves several key steps. Initially, a comprehensive dataset comprising both benign and ransomware samples is collected. This data undergoes preprocessing to clean and normalize it, followed by feature extraction to identify relevant attributes that can differentiate between normal and malicious activities. Various machine learning algorithms, such as Decision Trees, Random Forests, Naïve Bayes, and Neural Networks, are then employed to train models on the processed data.

Evaluation of these models is critical to ascertain their effectiveness in detecting ransomware. Metrics such as accuracy, precision, recall, and F-beta scores are utilized to compare the performance of different algorithms. Recent studies have shown that ensemble methods like Random Forest consistently outperform other classifiers in terms of accuracy and precision, making them particularly effective for ransomware detection tasks. Continuous research in this area aims to enhance these methodologies further, adapting to the evolving landscape of cyber threats and improving overall cybersecurity measures.

IV. SYSTEM ARCHITECTURE

The data is obtained through publicly available ransomeware files sourced from Kaggle. In the training phase first the sample set which includes various parameters were analysed then the extraction of the feature takes place after which the generation of the training feature set is assigned.

In the detection phase, the same procedure from analyzing to the extraction of the features are same, then the generation of the vector that helps in the detection of the trained set which can be seen in fig 4.1.

This allows seamless integration of the two phases of the model where the original asset of the project outcome is obtained which involves the detection of the trained dataset which results in the two different possible outcomes ie whether it is a ransomeware file or the safe file.

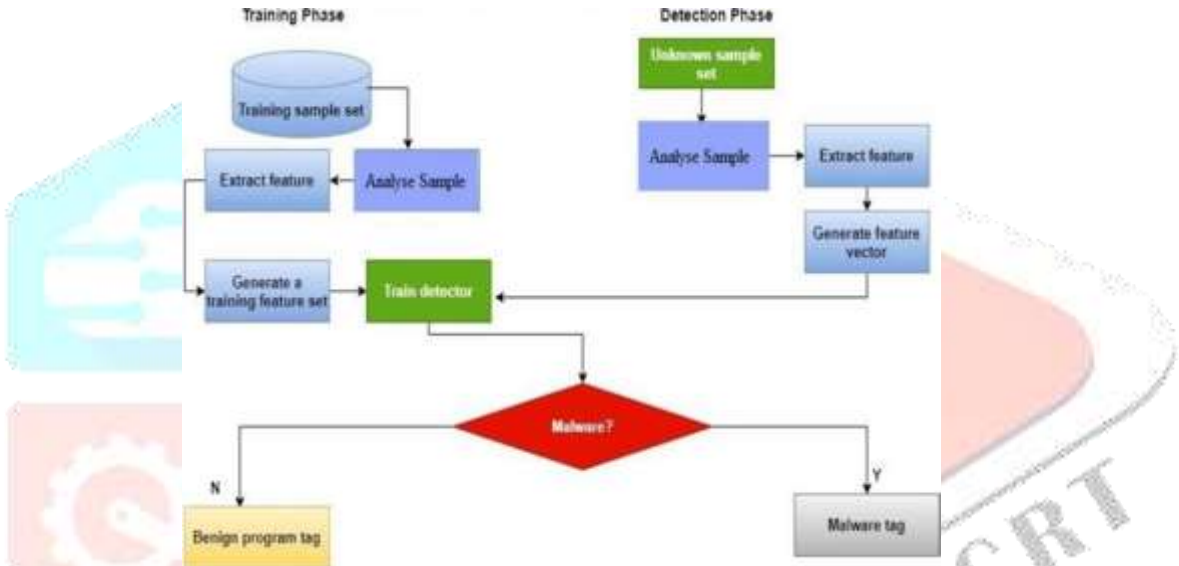

Figure 4.1:System Architecture

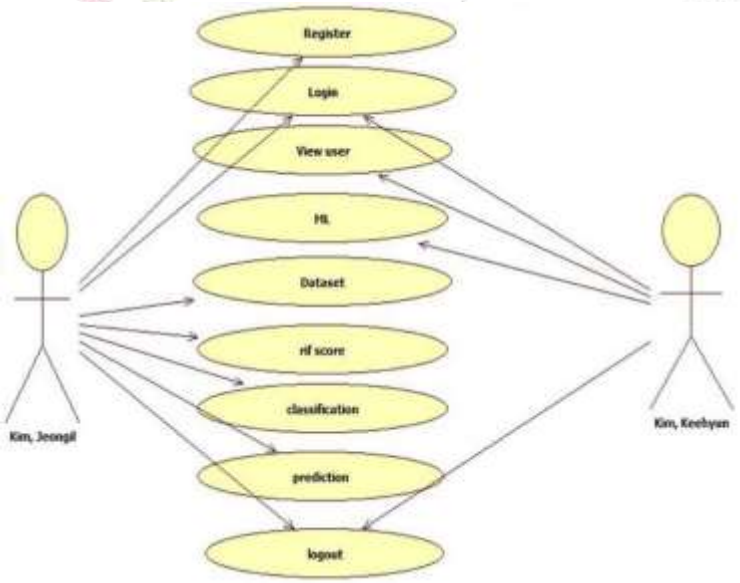

Figure 4.2:Use Case Diagram

V. CONCLUSION

Now a days ransomware has been at high point in cyber security. It is crucial to detect ransomware before it infects your machine. The paper provides information for detecting ransomware. Though various research papers are available many have drawbacks like accuracy. The goal is to detect ransomware before it infects machine with much accuracy. This paper has provided an overview of several different methods proposed by different researchers to addresses the challenge of Ransomware Detection. Many algorithms like SVM, Decision tree are available to detect ransomware. Here, in this paper Random Forest is used as it chooses best decision tree for detection of ransomware.

In conclusion, this review highlights the importance of Ransomware Detection. Detecting ransomware families is somehow challenge and according to many researches system fails here. Also this review helps in finding the various types of threats that can be occurred with respect to the data breaching can be seen and according to that the prevention form the ransomware attackers is to be considered.

References

[1] “Ransomware detection using machine learning: A review, research limitations and future directions” by JAMIL ISPAHANY1,2 (Graduate Student Member, IEEE), MD RAFIQUL ISLAM1,3 (Senior Member, IEEE), MD ZAHIDUL ISLAM1,2 (Senior Member, IEEE), and M. ARIF KHAN.1,2(Senior Member, IEEE).

[2] “Enhanced Ransomware Detection Techniques using Machine Learning Algorithms” by Mohammad Masum, Md Jobair Hossain Faruk, Hossain Shahriar, Kai Qian, Dan Lo, Muhaiminul Islam Adnan.

[3] “Ransomware detection using deep learning based unsupervised feature extraction and a cost sensitive Pareto Ensemble classifer” by Umme Zahoora, Asifullah Khan, Muttukrishnan Rajarajan*, Saddam Hussain Khan, MuhammadAsam & Tauseef Jamal.

[4] Benjamin Marais, Tony Quertier, Stéphane Morucci. AI-based Malware and Ransomware Detection Models. Conference on Artificial Intelligence for Defense, DGA Maîtrise de l’Information, Nov 2022, Rennes, France. ffhal-03881198f