Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Special Issue 4, April : 2024

DETECTION OF RANSOMWARE ATTACKS USING PROCESSOR AND DISK USAGE DATA

M.Ram Durga Apparao1,K.Likitha2,P.Uma Sri Lakshmi3,P.Deepthi4,K.Sravani5 1,2,3,4,5 Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women,Visakhapatnam,Andhra Pradesh,India

ABSTRACT
Ransomware often evades antivirus tools, encrypts files, and renders the target computer and its data unusable. The current approaches to detect such ransomware include monitoring processes, system calls, and file activities on the target system and analysing the data collected. Monitoring multiple processes has a very high overhead; newer ransomware may interfere with the monitoring and corrupt the collected data. This project presents a robust and practical approach to detecting ransomware in execution on a virtual machine (VM). This approach avoids the overhead of continuously monitoring every process on the target machine and prevents the risk of data contamination by ransomware. Furthermore, it is resilient to variations in user workloads. It provides fast detection with a high probability for known (used for training the ML model) and unknown (not used for training) ransomware. The random forest (RF), XG Boost and voting classifier performed the best of all the ML classifiers we tested. Over six different user loads and 22 ransomware. As an extension we used voting classifier for detection which help in getting relevant data for training and can improve detection accuracy.Also identified three different types of ransomware attacks crypto ransomware, locker ransomware, scareware.
Keywords:- Deep learning, disk statistics, hardware performance counters, machine learning, ransomware, virtual machines.

INTRODUCTION

Ransomware is malware that encrypts files on a target computer or locks the computer to render the target machine and its data unusable. Cyber attackers use ransomware attacks to extort victims' money. Nation-state actors may use ransomware attacks to inflict harm on the critical infrastructure of their adversaries. These attacks are often combined with the exfiltration of victims' data to compel the victim to pay a ransom or sell the data on the dark web. In 2022 around 70% of businesses were victimized by ransomware attacks . Ransomware is expected to attack a business, consumer, or device every 2 seconds by 2031, up from every 11 seconds in 2021. The damage cost was $20 billion in 2021 and is likely to exceed $265 Billion By 2031 . Several researchers have recently investigated the detection of ransomware attacks. Signature-based detection , relies on the hash values generated by antivirus software for known ransomware and checks the target machine for files that match the hash values. However, polymorphic and metamorphic versions of previously known ransomware can bypass such signature-based detection . Therefore, behavioral or runtime detection of ransomware in execution complements signature-based detection methods. The behavioral analysis is a dynamic analysis that looks into the virus's behavior—the sequence of actions performed by the ransomware after it infects the victim machine. While malware activities vary widely, ransomware must perform a specific sequence of activities to encrypt as many data files as possible quickly. Some recent ransomware such as LockBit2.0, Darkside, and Black Matter encrypt only parts (the first few bytes) of files to render more files unusable quickly . Therefore, the requirement of quickly encrypting user files is likely to differentiate ransomware's runtime behavior from that of a benign application. The premise is that a system under ransomware attack must exhibit some form of immutable anomalous behavior. For example, the ransomware must access files from the hard disk and use the processor for data encryption

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Special Issue 4, April : 2024

resulting in elevated activity; suitably trained machine-learning methods may detect the elevated activity.

LITERATURE SURVEY

1. On the classification of Microsoft-Windows ransomware using hardware profile (2021)Author :Sana Aurangzeb, Rao Naveed Bin Rais,Muhammad Aleem.
2. Know Abnormal, Find Evil: Frequent Pattern Mining for Ransomware Threat Hunting and Intelligence(2022)Author: Sajad Hashemi,Raouf Khayami Homayoun,Ali Dehghantanha,Marzieh Ahmadzadeh,Sattar.
3. RanStop : A Hardware-assisted Runtime Crypto-Ransomware Detection Technique (2020) Nitin Pundir,Mark Tehranipoor,Fahim Rahman.
4. RATAFIA: Ransomware Analysis using Time And Frequency Informed Autoencoders Behavior Based Ransomware Detection(2019)Author: Manaar Alam,Sarani Bhattacharya,Swastika Dutt,Sayan Sinha.

EXISTING SYSTEM

Runtime detection implemented on the target machine requires continual monitoring of various processes or components and subsystems, collecting data related to various events, and analyzing the data for anomalous behavior Ransomware may try camouflaging its runtime behavior by creating additional processes and activities. However, the fact remains that a system under attack exhibits some form of elevated activity, which is detectable with appropriate analysis Runtime detection is resource- intensive and intrusive if the monitoring is done on the target machine since the process corresponding to ransomware may not be easy to identify, and multiple processes must be monitored. Also, such monitoring is prone to be disabled by the ransomware designed to shut down active processes before encrypting files. If the monitoring is carried out on the target machine, runtime detection is resource-intensive and intrusive since it may be difficult to identify the ransomware-related process and multiple processes need to be watched over. It is also possible for ransomware that is meant to stop running processes before encrypting files to disable this kind of monitoring.

PROPOSED SYSTEM

In this we use different machine and deep learning models for analysis of the Ransomware Attack dataset As an extension we applied an ensemble method combining the predictions of multiple individual models to produce a more robust and accurate final prediction. However, we can further enhance the performance by exploring other techniques such as Voting Classifier, and got 99% of accuracy, We also designed the front end using the flask framework for user testing and with user authentication.

To overcome from above issue author of this paper employing VMWARE on host system which will read Hardware Performance Counters (HPC) and IO EVENTS data and then applying this data on machine learning models to predict whether executing script is normal (benign) or Ransomware. Extracting HPC and IOEVENTS features using VMware will not affect system performance and machine learning models also able to predict Ransomware with more than 90% accuracy. In propose paper author has experimented with various machine learning algorithms such as SVM, KNN, Decision Tree, Random Forest and XGBOOST. In all algorithms Random Forest, XGBOOST is giving accuracy.

Avoids the overhead of monitoring many processes on the target machine and prevents data contamination by the ransomware designed to thwart such monitoring activities. We evaluate the detection effectiveness under different user workloads. Furthermore, we demonstrate that machine

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Special Issue 4, April : 2024

learning (ML) and deep learning (DL) based detection models improve detection accuracy when trained with varying workloads. We demonstrate that combining the HPC and disk I/O data to build an ML- based detection model can improve the detection accuracy compared to using a model based on HPC or I/O data alone.

THE DESIGN STRUCTURE OF THE COMPARTATORS

Data loading: using this module we are going to import the dataset. Data Processing: Using the module we will explore the data.

Splitting data into train & test: using this module data will be divided into train & test

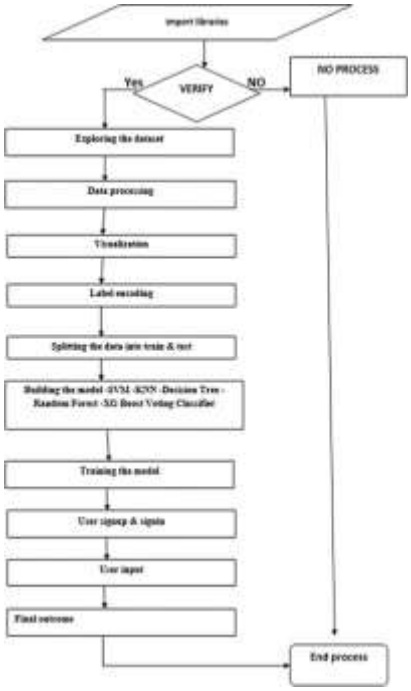
Model generation: Model building -SVM -KNN -Decision Tree -Random Forest -XGBoost-Voting Classifier. Algorithms accuracy calculated

User signup & login: Using this module will get registration and login User input: Using this module will give input for prediction

Prediction: final predicted displayed

Note: Extension: As an extension we applied an ensemble method combining the predictions of multiple individual models to produce a more robust and accurate final prediction. However, we can further enhance the performance by exploring other techniques such as Voting Classifier , and got 99% of accuracy.

As an extension we used voting classifier for detection which help in getting relevant data for training and can improve detection accuracy. Also identified three different types of ransomware attacks crypto ransomware, locker ransomware, scareware. We also build the front end using the flask framework for user testing and with user authentication.



RESULT ANALYSIS

This paper presents an approach to detect ransomware executing on a VM quickly and accurately by collecting processor and disk I/O activity events for the VM from the host machine and using machine learning techniques to analyze the data. The processor-event data are collected using the perf tool and

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Special Issue 4, April : 2024

hardware performance counters (HPCs) for five events, selected from more than 40 events using recursive feature elimination with cross-validation; disk I/O-event data are collected for eight events. We considered five ML. For each classifier, we developed three models: one uses HPC data only, the second uses disk I/O data only, and the third is an integrated model that uses both HPC and I/O data. The integrated model performed the best for all seven classifiers. The random forest (RF) classifier has the best detection accuracy among the seven classifiers, and its training times are lower than those of the other classifiers. Overall, the RF-integrated model shows promising results in detecting known ransomware (used for training) and unknown ransomware (not used in training).

Output Screens

Fig:Home Page

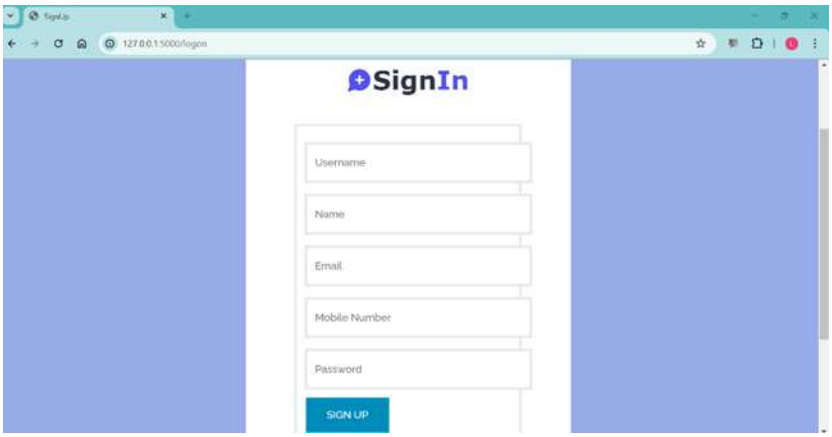


Fig:Sign Up Page

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Special Issue 4, April : 2024

Fig:Sign In Page

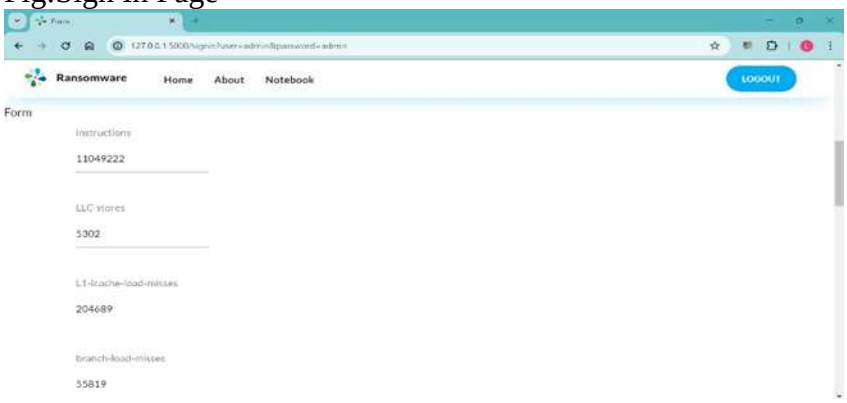


Fig:Form Page

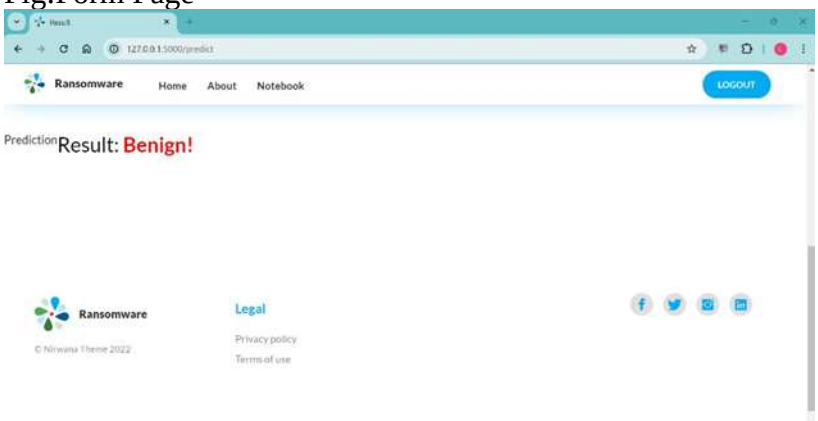


Fig:Benign Predicted Page

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Special Issue 4, April : 2024

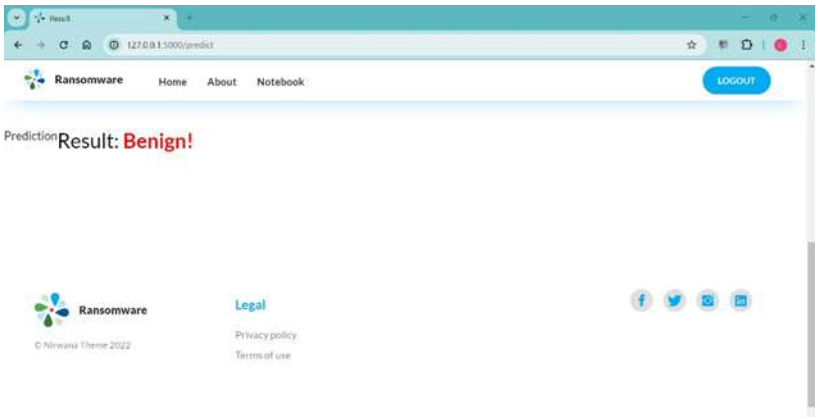


Fig:Crpto Ransomware Predicted Page

CONCLUSION

- Our project pioneers an efficient ransomware detection method, prioritizing accuracy and minimal performance impact.
- Rigorous evaluation highlight Random Forest, XGBOOST and Voting Classifier as top performing models for precise ransomware prediction.
- Sharing a public dataset forests collaboration, supporting advancements in ransomware detection research.
- Integrating with Flask and SQLite ensures both security and user accessibility, enhancing the practicality of our ransomware detection system.

REFERENCES

[1] SR Department. (2022). Ransomware victimization rate 2022. Accessed: Apr. 6, 2022. [Online]. Available: https://www.statista. com/statistics/204457/businesses-ransomware-attack-rate/

[2] D. Braue. (2022). Ransomware Damage Costs. Accessed: Sep. 16, 2022. [Online]. Available: https://cybersecurityventures.com/globalransomware-damage-costs-predicted-to-reach 250-billion- usd-by- 2031/ (2021).

[3] Logix Consulting. (2020). What is Signature Based Malware Detection. Accessed: Apr. 3, 2023. [Online]. Available: https://www.logixconsulting. com/2020/12/15/what-is-signature-based-malware-detection/

[4] Polymorphic Malware. Accessed: Apr. 3, 2023. [Online]. Available: https://www.thesslstore.com/blog/polymorphic-malware-andmetamorphic-malware-what- you-need-to-know/