Journal of Harbin Engineering University
ISSN: 1006-7043
Vol 44 No. 7
July 2023

An Early Detection Mechanism for Ransomware: A Survey

Srijita Bhattacharjee[1], Dr. Dhananjay Dakhane[2]

[1]RamraoAdik Institute of Technology, D.Y. Patil Deemed to be University, Nerul, Navi Mumbai, Maharashtra, India
[2]RamraoAdik Institute of Technology, D.Y. Patil Deemed to be University, Nerul, Navi Mumbai, Maharashtra, India

Abstract -Ransomware could be a frame of blackmail in which digital records are rendered blocked off until a ransom payment is made. The frequency of ransomware attacks has risen at a disturbing rate in recent years and the cybercriminals that deal with ransomware also have increased their skills. If victim chooses to pay the ransom amount, then also the encryption can cause the irreversible damage to the files. Protection from the ransomware attacks is a difficult job as there is less knowledge about the undetected ransomware or newly detected ransomware and ever evolving variants. So early phase detection is the key to secure the files from the encryption phase and demand ransom. Hence in this paper several proposed studies of early detection solutions have been explored and future research challenges and directions are also be highlighted.

Keywords-Ransomware, Evolving Variants, Encryption, Early Detection

1.Introduction

Malware known as ransomware encrypts a victim's data and holds it hostage until the victim pays a ransom. A client or organization's important data is scrambled so that files, databases, or applications are not accessed [1]. At that time, a ransom is demanded in order to grant access. Ransomware frequently targets database and file servers, spreads throughout a network, and can swiftly drag a whole corporation down. It could be a developing risk that causes significant financial harm to governments and commercial enterprises and payments to cybercriminals totalling billions of dollars in installments and payments [7].

Indeed, Despite the fact that ransomware has been making headlines for the past five years or so, it is not a novel idea to steal user data or encrypt files, prevent system access, and then demand a ransom to decrypt the data. Criminals held encrypted records hostage in return for cash sent via postal mail in the late 1980s [10]. One of the first known ransomware assaults was the AIDS trojan (PC Cyborg Virus), which was spread via floppy disc in 1989. Indeed, in spiteof the fact that it was a straightforward infection that utilized symmetric cryptography, Casualties required to send $189 to a P.O. box [24] in Panama to re-establish their access.Despite its long history, ransomware attacks remained limited till 2000s, due to the difficulties to gather the ransomware However, the rise of crypto currencies such as Bitcoin in 2010 altered [34] the entire payment collection process. Ransomware has turned into a profitable business by offering a simple and untraceable method of accepting payment from victims in the form of crypto currencies. The ground-breaking new ransomware family known as CryptoLocker not only commandeered Bitcoin transactions but also coupled them with more sophisticated encryption techniques [5]. It created 2048-bit RSA key pairs from a command-and-control server to encrypt the victim's files, making sure the victim couldn't escape until they paid for the key[6]. To propagate CryptoLocker, the banking Trojan Gameover Zeus was employed [36]. The FBI and CrowdStrike researchers worked together tobring CryptoLockerGameover Zeus to an end. Though it did not operate for the first seven months, it demonstrated to the entire cybercrime community the enormous business potential of ransomware. This was the real turning point in the quick evolution of ransomware. It was observed that cybercrime

increases pressure on victim organizations to pay their ransom by utilizing various methods. Data leak extortion techniques using ransomware are common among several eCrime groups in 2020, but data exfiltration by ransomware attacks were uncommon in previous years. When using this tactic, data from a victim is both encrypted and leaked[8] with the threat of doing so if the extortion demand is not met Cybercriminals that use ransomware have improved their skills and complexity over the past few years. It is reasonable to expect that this tendency will accelerate and that thieves will put as much pressure as possible on businesses to pay a big ransom.

Five Major ransomware attacks of 2022 are listed below [35]

1. **Nvidia:** In February, 2022 the world's biggest semiconductor chip company was attacked by ransomware.
2. **Costa Rica Government:** This has likely been the foremost spoken-of attack in 2022 as first time a nation declared a national crisis in reaction to a ransomware attack.
3. **Bernalillo County, New Mexico:**One of the biggest attacks in 2022 took place there in January. When several county departments and government offices were taken offline by a crippling ransomware attack, New Mexico discovered that it had become the victim.
4. **Toyota:** Three Toyota services were compromised between February and March 2022.
5. **SpiceJet:** Indian airline SpiceJet confronted a ransomware attack in earlier 2022, leaving hundreds of passengers stranded in several places within the country.

Ransomware attacks are on the rise, and you or your organisation could be the next victim. Ransomware protection is critical; regardless of your location or the nature of your business, Ransomware protection is required. The next phishing email may appear genuine to an employee of the organization, and this could be the beginning of the ultimate mess - compromised confidential data, encrypted files, blocked access, and more. As a result, no one is safe. Every association must prepare for the protection from ransomware attack else organization's reputation and money will be at stake.

The rest of this paper is structured as follows: in section 2 we have discuses about the types of ransomwares, Working methods and target of ransomware. Section 3 focuses on various approaches proposed by researchers to detect ransomware in the early phase and critical analysis of the surveyed literature. In section 4, we have presented Research challenges and future directions on early detection of ransomware. Finally, we conclude our review work in section 5.

2.Ransomware

Ransomware encrypts data before requesting a payment from the victim. In order to prevent a customer or company from accessing files, databases, or apps, vital data is scrambled. At that time, a ransom is demanded in order to grant access. Ransomware frequently targets database and file servers, spreads throughout a network, and can quickly bring down an entire company. can swiftly drag a whole corporation down. It can be a developing hazard that generates billions of dollars in payments and installments to criminals while causing significant financial harm to corporations and governmental organisations.

2.1 Type of Ransomware

Ransomware can be classified into three main forms [33]:

- Locker ransomware: This type of ransomware restricts essential computer operations. For example, the user will not be able to access the desktop, whereas the mouse and console are mostly impaired. Users are able to utilise the system while the window handles the ransom demand and collects the ransom.Otherwise, the computer is inoperable. But locker ransomware does not target the critical files usually it just wants to lock the access from the user. Total annihilation of the record is subsequently impossible.

● Crypto ransomware: This type of ransomware encrypts important and personal data, but does not to block the fundamental computer functions. Users become anxious because they can see their files but are unable to access them. Cryptographic designers frequently include a start-up ransom. In case a user fails to pay the ransom by the due date all the information will be erased. The majority of users are unaware of the importance of data backups in the cloud or on external physical storage devices, which might be destroyed by crypto-ransomware. Therefore, many victims would rather pay the ransom only to have their data back.

● Scareware ransomware: This sort of ransomware uses pop-up windows to trick users into thinking they must download some software, hence exploiting restrictive methods for malware downloads. Instead of blocking the device or encrypting the data, cybercriminals use scareware to capitalise on the panic.

2.2 Working Method of Ransomware

Ransomware disguise as a normal file which is downloaded by the user intentionally or unintentionally and enters devices as a Trojan. As soon as it executes, it starts encrypting the files on an infected device and usually displays a message for the hapless that if the attackers receive the ransom, then only files can be decrypted. If the user fails to pay the ransom within the deadline, then the ransom amount is increased or the encrypted files are deleted. When only the program developers have the access to the decryption key [9]then these ransomwares are the most effective and dangerous

2.3 Target for Ransomware

Attackers select the companies they wish to deploy ransomware on in a variety of ways. Sometimes it comes down to chance; for instance, because universities typically do not have sizable security teams, attackers may decide to target them. Attackers may choose [32] to target a certain company if they believe it will rapidly comply with a ransom demand. For instance, access to files is frequently needed quickly by organisations like the government or hospitals. In order to hide the fact that the data was hacked, law firms and other companies that deal with sensitive information would be willing to make a quick payment.

2.4 Do you have to pay the ransom?

Should you pay the ransom if the system has been compromised [36] and important data has been lost and cannot be recovered from a backup? Paying a ransom just motivates hackers to produce more malware. Attackers that use ransomware typically set a low ransom price that businesses can afford to pay quickly. Attackers frequently offer discounts for the early payment to entice victims to pay immediately before giving it much thought. This is done because some sophisticated malware alters the ransom price according to the country or region while detecting the country where the infected computer is running. Few businesses reserve Bitcoin specifically for ransom payments because, in general, the price point is not that high for businesses.

3.Related Work

Several research proposals have been made to find the ransomware.Presently; various aspects of ransomware are focused in studies, e.g., building different detection model using machine learning, behaviour characteristics-based detection, file-based detection.

Ricardo Misael et al, [1] collected fingerprints of the ransomware families using the pre-attack activities by utilizing more than 3000 samples. Techniques are used as Occurrence of Words in Natural Language Processing, API calls generated by ransomware evasion, machine learning algorithms and deep learning algorithms. Using these techniques, the classification of ransomware is done.S.H. Kok, Azween Abdullah, NZ Jhanjhi [2] developed a mechanism that provides two levels of pre-encryption detections of crypto-ransomware. This methodology has two stages, the first of which uses SHA-256 to compare the user's signature against known ransomware signatures. It provides early detection and fast and accurate detection is

achieved using this method. The learning algorithm models are used in the second level of detection which examine the API call generated during the preattack stage to detect both known and unknown ransomware[26].A framework for dynamic user-centric access control was proposed by Timothy McIntosha, A.S.M. Kayesa, Yi-Ping Phoebe Chen, Alex Ng, and Paul Watters [3] that gathers pertinent security indicators to make access control decisions that either indicates the UDAC or the CBI aspects of identity and behaviour of the application and the OS environmental factor.A prototype is used for an access control framework to recognise ransomware-like file system operations and to safeguard user data on Windows desktop platforms. The system developed a clear warning message to encourage users to make security-critical choices.The research on recent advancements in ransomware avoidance and detection was presented by Craig Beamana, Ashley Barkwortha, Toluwalope David Akandea, SaqibHakak a, Muhammad KhurramKhanb,[7], who also identified difficulties and directions for further research. they independently developed experimental ransomware using the analysis of a few well-known ransomware samples.AbdullahiArabo, Remi Dejoux, TimotheePoulain, and Gregoire Chevalier [13] explored the process behavior and its nature and checked the relationship between two things to determine if it is ransomware or not. The study suggests utilising machine learning to mimic the human immune system as a self-defence tactic. 34 malware samples, 41 pieces of benign software, and 7 ransomware samples were all examined. They differentiate between malicious programmes and ransomware, but with low false-positive and false-negative rates.For the initial stages of the attack, the Redundancy Coefficient Gradual Upweighting (RCGU) strategy was suggested by Bander Ali Saleh Al-rimy a et al.,[16]. That time the amount of data is insufficient. The technique was proposed to signify the crypto ransomware feature in a better way.The proposed technique was incorporated into the Mutual Information Features Selection's objective function toimprove the early detection of crypto-ransomware. Various types of threats like malware and intrusions can be detected at the early phase using this technique.

Bander Ali Saleh et al,[23]proposed a detection model where two techniques are implemented.Enhance semi-random subspace selection and incremental bagging,are two innovative strategies.Incremental bagging is utilized previously to capture the behaviour of ransomware in different attack phases. Enhanced semi random subspace selection was used for some features which can be optimal and noise free. Using these features, classifiers can be trained. A survey was taken for the best classifier and as per the majority classifier was chosen. The study presented a comparison of result between the proposed technique and the existing early detection solutions and showed a better result in terms of accuracy and data.Caio Carvalho Moreira et al,[42] presented the study's significant contribution of a static analysis method for ransomware detection. Data are extracted from PE header files and these data are used for ransomware detection. PE header files are classified using the Xception CNN algorithm, after converted into colour images in a sequential vector pattern.The performance of theXception-based ransomware detection method was assessed on two feature datasets and examined its advantages over rival models. One of these datasets was already developed with the data in accordance with the standards for classifying 25 significant ransomware groups. The tests showed that the suggested method successfully distinguishes between ransomware and legitimate software. Muhammad Shabbir Abbasi et al,[43] suggested an automationtechnique, in which features can be selected for ransomware detection and classification using the behaviour that use of particle swarm optimisation technique. The suggested method performs feature selection based on group significance and takes into account the importance of different feature groups.In the detection and classification of ransomware, feature groups of the data are used. The experimental results demonstrate that, in the majority of instances, the suggested method outperforms other cutting-edge methods used in this work for

benchmarking by a substantial margin. The importance of several feature groups and the characteristics picked by the suggested technique in ransomware identification and categorization are also presented in-depth study in this article.Jaskaran_Singh et al, [44] proposed a novel SINN based Ransomware Detection Scheme to present the systems for transforming the log files into new features and normalising the data. The proposed SINN-RD is secure against a number of potential attacks, according to the security study that was completed on it. The present log collection was used to compute an additional 22 features that assisted in the separation of ransomware from legitimate traffic. After processing, this expanded dataset contributed to making the model more resilient.TaranCyriac John et al, [45]effectively proposed a novel approach for the automatic evolution of a malice scorer.Furthermore, it was shown that the proposed CCMS model performed noticeably better than the most recent approach for automatically evolving malice scorers.The further investigation carried out in this study revealed that the malware families had an impact on the model's categorization determination.Monitoring network behaviour during an attack can help identify data exfiltration by double-extortion ransomware. The suggested strategy is adaptable enough to develop a different sub-model in line with network-related features, which might aid in identifying double-extortion ransomware.

3.1Early Detection Mechanism for Ransomware

Even after discovery and removal, ransomware, and particularly the effects of Crypto-ransomware, cannot be changed. To protect user data from encryption and force users to pay a ransom, early detection is therefore necessary.Several early detection solutions are proposed in different studies during the pre-attack phase. We have classified the research work based on different approaches are as follows:

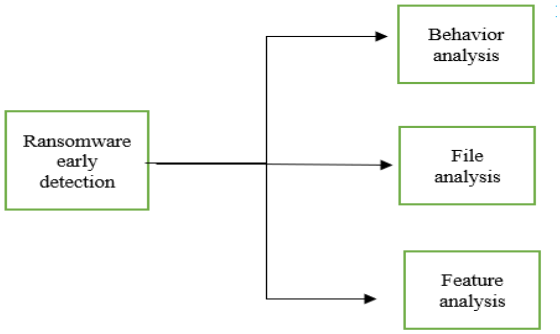

Figure 1: Classification of Methods for Early Detection

3.1.1 Behaviour Analysis

Various mechanisms are proposed by the researchers to stop ransomware assaults by utilizing various tactics, such as static/dynamic analysis approaches to determine the code structure of the virus and the activities which are done in its post infection behaviours [1,35]. Machine learning techniques can be incorporated to stop or detect those attacks. Though there is different previous work but the defence from the ransomware attack is very challenging due to the unawareness of different newly discovered ransomware and the families' ongoing evolution. In order to overcome the problems, we need to investigate several methods for detecting ransomware by examining the behavioural traits and assigning them to well-known harmful families. Ransomware families are programmed to execute some sequence or operation which can check the execution environment before the malicious execution. Recent detected ransomware families that have been found use calls to pre-attack APIs to recognise their surroundings and prevent execution, which is a detection evasion method [36]. Accomplish the first steps towards identifying ransomware can be done using these pre-attack activities, which can be categorized the ransomware families using these behavioural characteristics. Previous research [50] showed that the ransomware identification can be done using the API-based obfuscation techniques, in which malware developers purposefully alter the control flow of APIs to evade detection by anti-

malware software. Features are taken out of the logs that were gathered in the sandboxed environment. To find the best behavioural features for classification accuracy, [51]an iterative process is performed. The discovery of the behavioural characteristics that can be utilized to get the best accuracy and the machine learning algorithms are used to classify the ransomware. To address the issue of ransomware detection and multi-classification, a ransomware detection system is introduced. The research work named as RanPas [52] is a behaviour-based ransomware detection system which used the process tree and API calls as the foundation, Behavioural sequence of ransomware is also constructed in this.

3.1.2 File Analysis

Despite of the detection, we still face ransomware attack. So, we need to think on different perspective to halt this attack. Data from the user will be disturbed as soon as the machine is unable to identify the malware. To address the issue, detection of the files will be helpful rather than detection of executable program to set up the reinforcement framework just before the data are encrypted by the ransomware. The File based analysis strategy concentrate on the detection of the ransomware which shares similar file content or structural patterns [37]. After analysis the specific pattern of the ransomware, the detection rules can be applied to detect the malicious activity for the new ransomware. Ransomware can be detected and further activity can be prevented by identified malware activities [46].The algorithm is based on the evaluation of network probe-passively observed traffic. The programme was put to the test using 19 distinct ransomware families. According to the findings, it can spot ransomware activity before more than ten files are lost in less than 20 seconds. Because the content of those files was stored in the network probe's traffic logs, recovery of even those files was possible. A low percentage of false alarms were confirmed using several days' worth of traffic from actual corporate networks. A real-time solution for the early identification of crypto ransomware can be implemented using honey files.[47] The disc contains honey files that, if explored, will alarm the detection system, causing the ransomware operation to be stopped before encryption can start. Finding honey files' ideal location inside a file system is difficult. Current and prospective ransomware strains' file traversal patterns are examined by the system to decide where to distribute files across different directories.

3.1.3 Feature Analysis

Taking into account the data obtained during the pre-encryption stage, researchers have suggested various techniques for early detection.

Feature analysis is proved as a good option to identify the ransomware before it encrypts the files [38] But in the early stages of the attack, however, are data-limited. In the early stages of the attack, there wasn't enough information available, can has a negative impact on the feature selection techniques. As a result, the capacity of highlighting the characteristics of attack features can be reduced and it will decrease the performance of the detection model. To achieve superior redundancy relevancy trade-offs during feature selection, use the Redundancy Coefficient Gradual Upweighting (RCGU) technique. According to earlier research, the feature estimate technique hinges on comparing the candidate features and shared traits of the chosen features. Each feature in the chosen set as well as the potential feature has their shared data verified by the RCGU. The Mutual Information Feature Selection (MIFS) approach and the RCGU were therefore combined to create the Enhanced MIFS (EMIFS) [16]. As early detection depends on insufficient and incomplete attack patterns, which must be checked clearly, feature selection is crucial [38]. Despite the fact that this technique is effective when dealing with adequate attack patterns, it produces an incomplete feature set when dealing with complete data and observations about the attacks. [39]. The combination of different feature selection techniques had an impact on the classification precision rate [48] to evaluate the chosen column's application to classification. It was seen

that the input features gradually diminished. Not only does the accuracy show an improvement, but the decision tree's node count also goes down.
Finally, fewer rules improve intrusion detection performance, cut down on processing time, and enable early detection of abnormal traffic. About 8 feature selection algorithms were used to complete the feature selection. By comparing the outcomes of all feature filtration procedures, the simple majority vote process is used to choose the 19 crucial features [49]. The deep learning architecture demonstrates a strong ability to identify ransomware in the Android platform with greater accuracy and precision.

3.2 Critical Analysis

Table 1 provides a comprehensive evaluation and summary of the research on ransomware detection.

Table 1: Critical Analysis and Summary of Literature Survey

Paper citation No.	Methodology	Improvement	Limitation
1	Machine learning classification model	The dataset is quite big. Used confusion matrix. The speed of each model was computed to compare their performance.	Each family had a varied number of ransomware samples, So the dataset was not balanced. AV class was relied on to classify the samples into their family group.
2	Learning algorithm (RF)	By analysing the API before to encryption, the learning model may identify the ransomware using this method's early detection feature. This study discovered 14 significant APIs that can distinguish between ransomware and useful software.	The ransomware, which use a native encryption algorithm,is not detectable by the algorithm. This method is used for only one type of ransomware.
3	User-Driven Access Control (UDAC) and Content-Based Isolation (CBI)	Proposed a framework for access control that uses security indications to approve access requests The blocking of access requests that are neither intended nor authorised by the user is possible using UDAC and CBI principles.	More system security indicators can be added to systems to help identify fresh attack trends. It is necessary to recognise as many security indications as you can.
4	Decision tree, FRR/FAR	The block I/O packet header is used by the detection method to make a judgement. The system can identify distinctive I/O patterns that are seen while ransomware is active by keeping an eye on a stream of packets coming from the host.	When there is a high workload, the recovery queue grows enormously and there are more backup pages, which causes GC I/O overheads.The system cannot detect ransomware effectively.

5	Microsoft's Hyper-V	Efficiency of 78 decryption solutions developed by 11 security firms against 61 ransomware samples was tested. Evaluates the performance of the tools also	The sample contains only known ransomware and the number of samples is also less.
6	Hybrid Encryption Model	The reverse engineering process is a powerful tool to analyse ransomware so that effective decryption tools can be made.	Need to analyse the ransomware samples that belong to different categories.
7	Machine learning	Access control, key management, and data backups are examples of preventative measures that might reduce damage and possibly stop future assaults.	Detection was not adequate and the false-positive rate was high.
8	static analysis and dynamic analysis, key generation	The system of key generation, encryption allows better comprehend the ransomware's approach to attain their goals	Requires a defence approach.
9	Hooking system	When ransomware calls the functions attached to the libraries by the prevention system, the system stores the encryption key.	Need to take more sample of ransomware for the experiment
10	dynamic hooking techniques and asymmetric encryption	The system's effectiveness was evaluated against 20 successful families of real-world ransomware. Demonstrates that the system can restore all encrypted files from 12 of these families	Some current ransomware families quickly encrypt files after infection and grant access to the information for a brief period of time. The spyware destroys the decryption key after this period of time and demands a payment. All backups are currently useless for recovering from the infection.
11	Hybrid cryptosystem	Evolution of key management	Reverse encryption is very much challenging and time consuming.
12	Whitelist Based Ransomware Solution (WRS)	Malicious processes are blocked from accessing and encrypting files.	System cannot segregate the malicious process very effectively.
13	Machine learning	The solution's advantage is that it requires a dataset that includes both ransomware and non-ransomware. except a database of signatures.	Detection is not fast
14	entropy algorithm, machine learning	Lower erroneous detection rates significantly. It has a low GC overhead because it just manages the necessary backup pages.	Only identify ransomware by I/O access patterns.
15	Pattern mining algorithm,	Distinguish between ransomware and goodware samples and	The time to detect the ransomware can be reduced.

	Machine learning class	identify the families for the given ransomware at very less time.	
16	Redundancy Coefficient Mutual information, Gradual up-weighting Feature Choice	Gives a more accurate evaluation of the crypto-ransomware in the beginning of the attack lifecycle when there is less data.	When calculating feature significance using the RCGU technique, the conditional redundancy term was not taken into account.
17	DNAact-Ran using machine learning	The ML algorithm proved successful at detecting ransomware. The proposed DNAact-Ran method's effectiveness was confirmed using the real-time dataset.	The test provides some evidence that active learning classifiers are more effective at quickly identifying ransomware
18	Ransomware Analysis as Dialogue for Attribution and Reconnaissance	Contextual data from one attack can provide human analysts with a more idea of an adversary,	The algorithmic enhancements, data expansion, and system integration must be there in the system
19	Hybrid multilevel profiling	Due to its method of integrated analysis, which has a strong detection rate and very few false positives, the suggested framework demonstrates an efficient manner for security-related businesses and the research community.	Malware is constantly changing. As a result, dynamic analysis did not work properly with some samples. As a result, analysis was performed statically.
20	Dynamic Distributed Secure Storage	The encryption was quick and safe. presented a method for distributing data among several servers. Data can be redistributed to another node if any node is compromised.	This is a tactical choice as well as a purely technical one.
21	Support vector machine	Before determining the best entropy metric for feature values, analyse the encrypted file format practically.	Result must show for SVM poly
22	Machine learning based ensemble approach	Permissions and system call logs are the most crucial aspects, according to the research. These can identify and categorise ransomware and non-ransomware on Android.	The research should have addressed the characteristics of ransomware instances and feature extraction metrics can be provided to validate the performance of the secure defence.

3.3Comparisons of Results

In this paper comparative summary is shown by table and graph for the surveyed literature on ransomware early detection. The parameters used in the comparisons are the no. of ransomware samples, True Positive Rate (TPR), False Positive

Rate (FPR) and Accuracy of the detection techniques. Table 2 shows the results of surveyed literature and percentage allotment in different parameter and Fig.2 shows the comparison of the results of the surveyed literature.

Table 2: Results of the Ransomware Detection

Paper	True Positive Rate	No. of Ransomware Sample	No. of Ransomware families	False Positive Rate	Accuracy
S.H. Kok et al (2020) [2]	___	995	11	___	99%
Ricardo Misael Ayala Molina et al (2021) [1]	___	3000	5	___	94.92%
Bander Ali Saleh Al-rimy et al (2021) [16]	96.21~99.40	39,378	15	03.84~00.68	95.73%~99.09%
Firoz Khan et al(2020)[17]	87.90%	426	11	10%	87.90%
Chia Ming Hsu et al (2021) [21]	___	4 kinds	___	___	92%
Bander Ali Saleh Al-rimy (2019) [23]	92.80%	8152	15	7.15%	92.30%
Shaukat (2018) [40]	98.25%	574	12	0.56%	___
Kharraz et al(2017)[25]	99.90%	504	12	5.90%	___
Takeuchi et al(2018)[27]	96%-99%	38152	5	2.40%	99.30%
Lee et al(2019)[31]	99.4%-100%	___	___	___	99.70%
Daniel Morato et al(2018)[46]	___	50	19	___	99%
Yahye Abukar Ahmed et al(2020)[53]	___	1354	14	41.30%	58%-78%

*__ means the information was not found in the Literature survey.

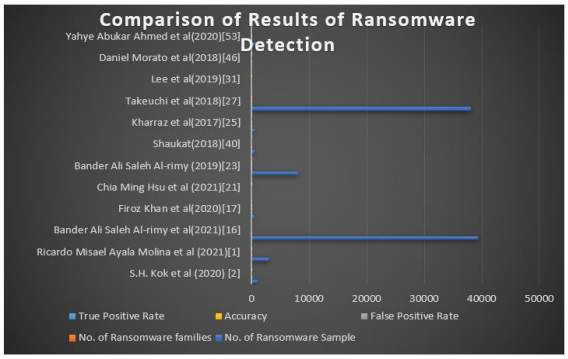

Figure 2 Comparison the Results of Ransomware Detection.

The primary issue in the detecting industry is accuracy. A powerful detection system will quickly and accurately identify ransomware. A novel adaptive pre-encryption ransomware early detection system which can addresses the population drift idea and limitations of previous research will be an original piece of work. If the system is not able to distinguish between genuine, benign and malicious one then the system will have a low accuracy. The accuracy will be less if a model is unable to recognize zero-day assaults or deal with how -ransomware attacks have grown and changed over time.

Several well-known assessment indicators, including Accuracy, True Positive Rate (TPR), False Positive Rate (FPR), No. of Ransomware samples and no of ransomware family is used to measure and compare the detection performance of reviewed literatures. TPR refers to the proportion of positive samples, such as ransomware, that are accurately identified. FPR, on the other hand, measures the proportion of incorrectly identified negative cases. The overall level of accuracy is calculated as the ratio of correctly identified samples, whether positive or negative, divided by the total samples.

Ransomware is detected in pre-encryption stage using eleven API calls [2]. Ransomware samples and good ware samples are collected from many sources and both can be detected using APIs.After the sample collection, cuckoo sandbox is used to provide the API call list as shown in a sample. A prediction model to identify the ransomware is created while training the data using Random Forest method. Once the ransomware is detected, it's signature and other important details are stored in the database. This process is significantly faster, safer and accurate. The malware, which uses its personal native encryption code, cannot be identified by this process. So, the method can be used as a supplement not as a solely technology to detect ransomware. The system can identify only one type of ransomware but the fact is there are more than one type of ransomware is available to hamper the business. Here cuckoo sandbox is used to generate the report of analysis of the samples but the installation part of the software is quite difficult so the research work needs to have one independent standalone application without the need of separate software. 23 Evasion API calls are invoked to get the behavioral characteristics [1] of ransomware samples and these behavioral characteristics can classify the ransomware into different ransomware families.19,499 samplers are collected from different sources e.g, VirusShare, VirustTotal and etc. The significantly large dataset is collected for ransomware samples but for testing the model, about 3000 samples are taken, five ransomware families are included in the samples at random, along with one benign family. Cuckoo sandbox is used for dynamic analysis. Different classification models are used to classify ransomware into families. Random Forest algorithm has given the highest accuracy among the 6 classification models. Pre attack activities of ransomware are captured to provide the classification of ransomware family. The work cannot be ideal for unseen or novel variant of ransomware as it used supervised learning. So unsupervised learning can be incorporated to get the desired result. In this research dynamic analysis is used but if static and dynamic both analyses are provided then the advantages of both analyses can be added in the system. Using the information gathered prior to encryption, the feature extraction is done efficiently by the system [16]. Early detection is captured using the API calls which are the functions in the victim's operating system and these are relevant to the encryption process. Dynamic analysis is done by cuckoo sandbox and after that feature selection process is done using the novel method and then the model is trained using different machine learning algorithms and it shows in this research that random forest is providing better accuracy. At early stage in the selection process, feature relevancy is important. In the early stages of the selection process, there were fewer features in the selected set, which decreased the likelihood that the candidate feature would be redundant. The information or data is insufficient prior to encryption. To address the issue, the research implemented the feature extraction method for ransomware at pre attack phase with less available

data. Various Ransomware detection methods exist, e.g signature-based detection, Hashing, and Machine Learning. Each strategy has benefits and drawbacks. The signature-based, technique identify ransomware by matching the patterns and observing the APIs. Butsignature-based technique will fail without an updated signature if ransomware changes its behavior or disguise its malicious code. DNA act Ran method [17] uses machine learning to detect ransomware using the digital DNA sequence. At first, sufficient feature is extracted using feature extraction techniques then the DNA sequence is generated for those features and finally it can classify between the ransomware and good ware. As soon as the system is unable to identify the ransomware, user data will be compromised. To solve the problem, detection of files[21] may be useful rather than the executable programme. A backup system should be set up before ransomware encrypts user files. Crypto ransomware encrypts files using diverse methods, to boost the rate of Crypto ransomware detection, encrypted file formats are first analysed. Next, the optimal entropy measure is evaluated as feature values, and a support vector machine model is implemented for it. The study [23] was introduced to create a detection model with enough knowledge about ransomware behavior during the pre-attack stages. To meet the requirement of the research, the technique was proposed to generate numerous data subsets to imitate the attack's evolution through the various phases. Furthermore, the technique excludes weak features while maintaining some diversity among the different subspaces to improve detection accuracy. The study [40] focuses on layer-based ransomware defence. Based on a huge dataset of Ransomware families, it generates a set of features that characterise Ransomware behavior using both static and dynamic analysis. Machine learning is used to identify the attack at early stage. When the first layers identify a process as having suspected Ransomware behavior, Then the files modified by the process are backed up to protect user data until the process is classed as Ransomware or goodware. The study offered a method [25] to maintain a transparent buffer for each storage I/O with relatively little operating system modification. All process's I/O request patterns are examined by the system for signs of ransomware-like behavior. In the event that I/O request patterns are found, as ransomware like then the process need to be killed and the data restored. A ransomware detection scheme is proposed using SVM model [27]. The system uses supervised learning. The motivation behind the idea is to train the model using the API calls of ransomware as its features, so that the system can detect previously undiscovered ransomware. But for that supervised learning is not sufficient. The research needs to have unsupervised learning to get the desired result. Ransomware is detected using the file entropy analysis [31]. The detection mechanism used machine learning to identify the infected files using the entropy analysis. A new approach is presented [46] to identify the new ransomware families by adding an acceptable no. pf randomly selected benign samples in the data set for the testing. The superior approach can be used as a good ransomware detection strategy. System calls, which might be an interface for programmes to seek services from the operating system, are the fundamental building blocks of the dynamic analysis. Anti-ransomware technologies' ability to detect high-noise behavioral sequences, which can be caused by redundant and pointless system calls, is diminished. A non-signature-based detection methodology [53] is developed using Windows API call sequences. The noisy data is removed while the most pertinent elements are chosen to describe the ransomware's activities.

Accuracy is dependent on the total sample taken for the experiment. We can observe that though there is various research work on pre-attack ransomware detection but different methods and models have different accuracy. Mostly researchers have explored the machine learning algorithm to train the model and got their desired result but the detection and experiments are dependent on the samples which they have taken from different resources. Reviewed literature shows that machine learning is very useful to detect the ransomware and there is very less

success for detection of the new or unseen ransomware.

4.Research Challenges and Future Research Directions

The research problems based on the reviewed literature are addressed in this section, and we have looked into potential future study topics.The user's unawareness, the False Positive Rate (FPR) for ransomware, the reliance on the Windows API, and the lack of a classification system for newly discovered or undiscovered ransomware are among the research challenges that have been examined.Edge and Fog based technologies, are among the Future research priorities to identify and prevent ransomware. Future research focuses on Deepfake ransomware, and Blockchain-based methods to detect ransomware.

4.1Research Challenges

User's awareness is one of the major aspects that can reduce the impact of ransomware attack. Though, no full proofed framework is there, which is able to reliably counter the ransomware attack which propagates the ransomware through the phishing campaign but user's knowledge to understand the spam email, which can be malicious can save their system from the infection.

Early detection of ransomware can mitigate the danger of encryption of user's data.While detecting the attack in the early stage, false positives are an important factor that needs to be remembered as existing study shows a limitation for the same. System will be ineffective if there is a low detection rate where as high False Positive rate can confuse the admin.

Another major restriction is ransomware's reliance on the Windows API, [2] which offers a more effective encryption technique. Another drawback is that an assault that employs its own native encryption code cannot be discovered.Defending against ransomware attacks or mitigating them requires identifying the ransomware earlier to execute their payload, which results in the encryption or blocking the focus on the device or information. An approach to achieving this is frequently to examine the behavioural traits of suspicious apps and identify adversarial behaviours that the proper defence mechanisms are triggered by and that could perhaps be ascribed to well-known ransomware groups [1]. Existing research, however, is unable to categorise novel, previously unknown ransomware varieties. The sustainability or weakening problem that most learning-based finders face is not addressed by existing research. The problem is created when existing samples are used to find new examples, and this affects how strong the model is.

Device-level backup, Solid State Drives have been proposed to protect against ransomware attacks. With this, only I/O access [4] patterns can be used to identify ransomware. However, if an assault does not display the typical I/O patterns of ransomware, it is usually not discovered.

4.2 Future Research Directions

The topic of edge and fog-based technologies offers a wealth of research options. These technologies can be used efficiently to detect ransomware attack [28]. Machine learning approaches can be used in this field to detect or prevent the ransomware. Federated learning can be incorporated in Edge and Fog based technologies to get the desired outcome.

Deepfake, is a threat utilizing artificial intelligence/machine learning (AI/ML) to formconvincing, realistic videos, images, and audios, which never happened. In future Attacker can design ransomware which can generate automatically deepfake contents such as videos or images of the user and will make the user to pay the ransom amount to restrict the publication of fake content online.

Blockchain has a decentralized nature along with the hash function linked with it and timestamp function. Tampering the user data is difficult due to this decentralized nature of Blockchain [30]. The future research directions can be the exploration of blockchain based technologies and make a block chain-based solution to mitigate the ransomware attack.

5.Conclusion

Defending against ransomware attacks is considered a difficult task because of the less information about the new or unknown ransomware and the constantly evolving variants. These attacks have been predominant so it is crucial to design solutions that can defend the system against ransomware attacks. So effective approaches need to be designed for early detection So before the actual attack or in the pre-encryption phase, ransomware can be detected and user can safeguard the system. In this paper, early detection strategies of ransomware were explored. It was found that the focus of early detection techniques mostly revolves around behavior analysis, file analysis and feature analysis approaches. Though, many researchers used machine learning based mechanisms to detect the ransomware, but it seems that there is a scope to use more intelligent mechanisms to identify the ransomware in early phase of attack. Finally,we highlighted the existing research challenges in the field of ransomware and listed future research directions.

References

[1] Ricardo Misael Ayala Molina , SadeghTorabi†, Khaled Sarieddine , Elias Bou-Harb, Nizar Bouguila , and Chadi Assi"On Ransomware Family Attribution Using Pre-Attack Paranoia Activities" IEEE Transaction 2021

[2] S.H. Kok, Azween Abdullah, NZ Jhanjhi "Early detection of crypto-ransomware using pre-encryption detection algorithm" Elsevier 2020

[3] Timothy McIntosh,∗ A.S.M. Kayes a, Yi-Ping Phoebe Chena, Alex Ng a Paul Watters a,"Dynamic user-centric access control for detection of ransomware attacks" Elsevier 2021

[4] SunghaBaek, Youngdon Jung, David Mohaisen, Senior Member, IEEE, Sungjin Lee, and DaeHunNyang," SSD-Assisted Ransomware Detection and Data Recovery Techniques" IEEE Transaction 2021

[5] BurakFiliz , Budi Arief , OrcunCetina, Julio Hernández-Castro," On the Effectiveness of Ransomware Decryption Tools" Elsevier 2021

[6] FabrizioCicala, Elisa Bertino" Analysis of Encryption Key Generation in Modern Crypto-Ransomware" IEEE Transaction 2020.

[7] Craig Beamana, Ashley Barkwortha, Toluwalope David Akande, SaqibHakak , Muhammad Khurram Khan, "Ransomware: Recent advances, analysis, challenges and future research directions" Elsevier 2021

[8] PranshuBajpai, RichardEnbody "Dissecting .NET ransomware: key generation, encryption, and operation " Elsevier 2020

[9] Lee et al., "Ransomware prevention techniques using key backup," LNICST 194, pp. 105–114, 2017

[10] Kolodenker E, Koch W, Stringhini G, Egele M. Pay break: Defense against cryptographic ransomware. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security; 2017. p. 599–611.

[11] Bajpai P, Sood AK, Enbody R. A key-management-based taxonomy for ransomware. APWG Symposium on Electronic Crime Research (eCrime); 2018. p. 1–12

[12] Kim D, Lee J. "Blacklist vs. whitelist-based ransomware solutions". IEEE Consum. Electron. Mag. 2020

[13] AbdullahiArabo, Remi Dejoux, TimotheePoulain, Gregoire Chevalier "Detecting Ransomware Using Process Behavior Analysis " Elsevier 2020

[14] Donghyun Min, Yung-woo Ko, Ryan Walker, Junghee Lee, and Youngjae Kim, A Content-based Ransomware Detection and Backup Solid-State Drive for Ransomware Defense " IEEE Transaction 2021

[15] Sajad Homayoun , Ali Dehghantanha , "Know Abnormal, Find Evil: Frequent Pattern Mining for Ransomware Threat Hunting and Intelligence" IEEE Transaction 2020.

[16] Bander Ali Saleh Al-rimy a et al., "Redundancy Coefficient Gradual Up-weighting-based Mutual Information Feature Selection technique for Crypto-ransomware early detection" Elsevier 2021

[17] FIROZ KHAN et al., "A Digital DNA Sequencing Engine for Ransomware Detection Using Machine Learning" IEEE access 2020.

[18] H. Van Dyke Parunak et al., "A Grammar-Based Behavioral Distance Measure Between Ransomware Variants" IEEE Transaction 2021.

[19] UBASH POUDYAL et al., "Analysis of Crypto-Ransomware Using ML-Based Multi-Level Profiling" IEEE access 2021.

[20] Jason Castiglione et al., "Dynamic Distributed Secure Storage Against Ransomware" IEEE Transaction 2020.

[21] CHIA-MING HSU et al., "Enhancing File Entropy Analysis to Improve Machine Learning Detection Rate of Ransomware" IEEE access 2021.

[22] Usman Ahmed et al., "Mitigating adversarial evasion attacks of ransomware using ensemble learning" Elsevier 2022

[23] Bander Ali Saleh et al "Crypto-ransomware early detection model using novel incremental bagging with enhanced semi-random subspace selection" Elsevier 2019

[24] Z. Cohen and G. Sands, "Four key takeaways on the USgovernmentresponsetothepipelineransomwareattack"May2021.[Online].Available:https://www.cnn.com/2021/05/11/politics/colonialpipeline-cyber hearing-senatehomelandsecuritycommittee/index.html

[25] Amin Kharraz, Engin Kirda, Redemption: Real-Time Protection Against Ransomware at End-Hosts, International Symposium on Research in Attacks, Intrusions, and Defenses RAID 2017: Research in Attacks, Intrusions, and Defenses pp 98–119

[26] A. AlSabeh et al., "Exploiting ransomware paranoia for execution prevention," in IEEE Int. Conf. on Communications (ICC). IEEE, 2020, pp. 1–6.

[27] Takeuchi et al, Detecting Ransomware using support Vector Machines, 47th International Conference on Parallel Processing,2018

[28] Mukherjee M, Shu L, Wang D. Survey of fog computing: fundamental, network applications, and research challenges. IEEE Communications Surveys & Tutorials 2018;20(3):1826–57.

[29] Zhang et al., "Ransomware classification using patch based cnn and self-attention network on embedded ngrams of opcodes," Future Generation Computer Systems, vol. 110, pp. 708–720, 2020

[30] Hakak S, WZ Khan WZ, Gilkar GA, Haider N, Imran M, Alkatheiri MS. Industrial wastewater management using blockchain technology: architecture, requirements, and future directions. IEEE Internet of Things Magazine 2020;3(2):38–43

[31] Lee K, Lee S, Yim K. Machine learning based file entropy analysis for ransomware detection in backup systems. IEEE Access 2019;7:110205–15

[32] possible impact of Ransomware | Information Security Office (berkeley.edu)https://security.berkeley.edu/faq/ransomware/what-possible-impact-ransomware.Accessed March 2023

[33] Kaspersky, "What are the different types of ransomware?" Dec 2020. [Online].Available:https://www.kaspersky.com/resourcecenter/threats/ransomware examples. Accessed March 2023

[34] A Brief History of Ransomwarehttps://www.crowdstrike.com/cybersecurity101/ransomware/history-of ransomware/#:~:text=One%20of%20the%20first%20ransomware,virus%20that%20utilized%20symmetric%20cryptography. Accessed Feb 2023

[35] 5 Major Ransomware Attacks of 2022-https://www.cm-alliance.com/cybersecurity-blog/5-major-ransomware-attacks-of-2022, Accessed Feb 2023

[36] Ransomware Payments – Should Companies Pay Or Not? https://www.gartner.com/en/articles/when-it-comes-to-ransomware-should-your-company-pay,Accessed Feb 2023

[37] E. Berrueta et al., "A survey on detection techniques or cryptographic ransomware," IEEE Access, vol. 7, pp. 144 925–144 944, 2019.

[38] H. Zhang et al., "Classification of ransomware families with machine learning based on n-

gram of opcodes,"Future Generation Computer Systems, vol. 90, pp. 211–221, 2019.

[39] U. Urooj, M. A. B. Maarof, and B. A. S. Rimy, ''A proposed adaptive pre-encryption crypto-ransomware early detection model,'' in Proc. 3rd Int.Cyber Resilience Conf. (CRC), Jan. 2021, pp. 1–6.

[40] Saiyed Kashif Shaukat; Vinay J. Ribeiro, RansomWall: A layered defense system against cryptographic ransomware attacks using machine learning, 2018 10th International Conference on Communication Systems & Networks (COMSNETS)

[41] J. Che, Y. Yang, L. Li, X. Bai, S. Zhang, C. Deng, Maximum relevance minimum common redundancy feature selection for nonlinear data, Inform. Sci. 409 201768–86, http://dx.doi.org/10.1016/j.ins.2017.05.013.

[42] Caio C. Moreira , Davi C. Moreira , Claudomiro de S. de Sales Jr., Improving ransomware detection based on portable executable header using xception convolutional neural network , Elsevier 2023

[43] MuhammadShabbir Abbasi , Harith AlSahaf , Masood Mansoori , Ian Welch, Behavior-based ransomware classification: A particle swarm optimization wrapper-based approach for feature selection, Elsevier May ,2022

[44] Jaskaran Singh, Keshav Sharma, Ashok Kumar Das, SINN-RD: Splineinterpolation-envisioned neural networkbased ransomware detection scheme, Elsevier January,2023

[45] TaranCyriac John , Muhammad Shabbir Abbasi , Harith AlSahaf , Ian Welch , Ju lian Jang-Jaccard , Evolving malice scoring models for ransomware detection: An automated approach by utilising genetic programming and cooperative coevolution, Elsevier ,June 2023

[46] Daniel Morato , Eduardo Berrueta , Eduardo M agaña , Mikel Izal, Ransomware early detection by the analysis of file sharing traffic, Elsevier December 2018

[47] Shina Sheen , K Asmitha , Sridhar Venkatesan ,R-Sentry: Deception based ransomware detection using file access patterns, Elsevier , October 2022

[48] Yu-Lun Wan, Jen-Chun Chang, Rong-Jaye Chen, Shiuh-Jeng Wang, Feature-Selection-Based Ransomware Detection with Machine Learning of Data Analysis, 2018 3rd International Conference on Computer and Communication Systems

[49] Iram Bibi , Adnan Akhunzada Jahanzaib Malik , Ghufran Ahmed and Mohsin Raza, An Effective Android Ransomware Detection Through Multi-Factor Feature Filtration and Recurrent Neural Network, 2019 UK/ China Emerging Technologies (UCET)

[50] May Almousa, Sai Basavaraju , and Mohd Anwar, API-Based Ransomware Detection Using Machine Learning-Based Threat Detection Models, IEEE,2021

[51] Hajredin Daku, Pavol Zavarsky, Yasir Malik, Behavioral-Based Classification and Identification of Ransomware Variants Using Machine Learning, 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications

[52] Bidong Wang, Hui Liu, Xinli Han, Dongliang Xuan, RanPAS: A Behavior-based System for Ransomware Detection, 2021 IEEE Sixth International Conference on Data Science in Cyberspace (DSC)

[53] Yahye Abukar Ahmed , Barıs¸ Koçer , Shamsul Huda , Bander Ali Saleh Al-rimy , Mohammad Mehedi Hassan, A system call refinement-basedenhanced Minimum Redundancy Maximum Relevance method for ransomware early detection, Elsevier,2020

[54] Kumar, S. A. S., Naveen, R., Dhabliya, D., Shankar, B. M., & Rajesh, B. N. (2020). Electronic currency note sterilizer machine. Paper presented at the Materials Today: Proceedings, , 37(Part 2) 1442-1444. doi:10.1016/j.matpr.2020.07.064 Retrieved from www.scopus.com

[55] Kumbhkar, M., Shukla, P., Singh, Y., Sangia, R. A., & Dhabliya, D. (2023). Dimensional reduction method based on big data techniques for large scale data. Paper

presented at the 2023 IEEE International Conference on Integrated Circuits and Communication Systems, ICICACS 2023, doi:10.1109/ICICACS57338.2023.10100261 Retrieved from www.scopus.com

[56] Mandal, D., Shukla, A., Ghosh, A., Gupta, A., & Dhabliya, D. (2022). Molecular dynamics simulation for serial and parallel computation using leaf frog algorithm. Paper presented at the PDGC 2022 - 2022 7th International Conference on Parallel, Distributed and Grid Computing, 552-557. doi:10.1109/PDGC56933.2022.10053161 Retrieved from www.scopus.com