AI-enhanced Defense Against Ransomware Within the Organization's Architecture

B. N. Chaithanya* and S. H. Brahmananda

CSE, GITAM School of Technology, Bangalore, India
E-mail: cnagaraj@gitam.edu; bsavadat@gitam.edu
**Corresponding Author*

Received 27 December 2021; Accepted 09 August 2022;
Publication 07 November 2022

Abstract

Ransomware is a type of revenue-generating tactic that cybercriminals utilize to improve their income. Businesses have spent billions of dollars recovering control of their resources, which may include confidential data, operational applications and models, financial transactions, and other information, as a result of malicious software. Ransomware can infiltrate a resource or device and restrict the owner from accessing or utilizing it. There are various obstacles that a business must overcome in order to avoid ransomware attacks. Traditional ransomware detection systems employ a static detection method in which a finite dataset is provided into the system and a logical check is performed to prevent ransomware attacks against the system. This was effective in the early stages of the internet, but the scenario of recent times is far more advanced, and as more and more cyber world contrivances have been analyzed, multiple gaps have been identified, to the benefit of ransomware attackers, who use these gaps to generate astronomically large sums of money. As a result, the suggested methodology aims to efficiently detect diverse patterns associated with various file formats by starting with their sources, data collecting, probabilistic identification of target devices,

Journal of Cyber Security and Mobility, Vol. 11_4, 621–654.
doi: 10.13052/jcsm2245-1439.1146

and deep learning classifier with intelligent detection. An organization can use the recommended approach to safeguard its data and prepare for future ransomware attacks by using it as a roadmap to lead them through their security efforts.

Keywords: Ransomware detection, ransomware prediction, data analysis, natural language processing, deep learning, LSTM.

1 Introduction

Since the birth of the cyber world and subsequent advancements, cybersecurity has become a rising concern. By making information more available and connecting individuals, networks, and systems, the internet has altered information and communication management. This ease of access has resulted in unlawful operations, data theft, and damage implantation in order to generate cash through cyber-attacks. This unauthorized access to information is performed via cyber-attacks, which have increased in popularity as a technique of illicit money creation and have created significant hurdles to businesses and end users [1].

Unauthorized or malicious access to computer resources and privileged access data should be avoided at all costs, including the deployment of malware to seize control of computer resources and privileged access data, according to the International Cyber Security Alliance. Many malicious programs have been found in the most recent evolutionary browsers, all of them are intended to cause damage to confidential information [2]. In one type of a ransomware attack, a file is installed on a computer that gives access to critical digitalized resources but subsequently blocks access until the ransom is paid. When a ransomware attack is launched, the strategic strategy begins with the insertion of a file that grants access to vital digitalized resources, after which access is restricted until the ransom has been paid.

Cybersecurity threat imposers plan their attacks by researching potential components and disseminating harmful ransomware on the user/resource demeanor. Ransomware attackers adapt and invent ways based on previous behavior, taking advantage of technology advancements. These attackers collect the most information and deliver attacks to the weakest points in the source systems, apps, networks, and so on. Previously, preventive measures were offered. Still, ransomware dominates all aspects of the cyber world's resources (e.g., PCs, workstations, servers, payment connections, web links, and so on) (E.g., PC, workstations, servers, Payment links, web links, and

more). Traditional methods fail to detect the infection because they are rather static and lack learning abilities. The frequency of ransomware attacks has increased in recent years; the increased number of online users directly correlates to an increase in the number of attacks. In order to provide a safe ecosystem for the cyber-secure world's operation, emerging and critical technologies must be put into action.

Ransomware is a form of computer virus that encrypts data and locks the offender's screen in order to demand a ransom and wreak significant damage. The time-consuming task of dynamic analysis can be accomplished by ransomware that is totally aware of its surroundings. Ransomware is becoming increasingly frequent, posing a threat to internet users, governments, and enterprises all around the world. Ransomware is one of the most common types of malware. It encrypts a user's sensitive data and only unlocks it if a ransom is paid. As malware makers shift their product distribution from HTTP to HTTPS to evade payload analysis, there will be no longer the ability rely on deep packet inspection to extract functionality for malware detection of data which monitors network traffic between an infected PC, command and control server. Because of the obvious ransomware-as-a-service concept, which makes it exceedingly easy to obtain and use, as well as the possibility for massive profits, ransomware has become a sustainable criminal business plan. Ransomware attacks can harm consumers, private businesses, and public sector organizations such as hospitals and utilities providers, causing substantial disruption and financial loss. Despite the advancement of machine learning approaches for detecting ransomware, new variations are being developed to circumvent detection when dynamic machines are used.

Ransomware has risen to the top of the list of all malware as a result of traditional malware attacks [3]; ransomware is not only causing damage to digital ecosystems, but it is also causing human deaths due to data mishandling and the erroneous implementation of machines that have been triggered and deployed by the attacker. According to Khayami [4], the first ransomware-related human death occurred in Germany in 2020. Ransomware attacks users, government sectors, small enterprises, and multinationals on an enormous scale. As a result of this practice, businesses are losing millions of dollars. Traditional defense management systems are not designed with the various targeted platforms in mind, making these platforms vulnerable to attack as a result. The major targeted platforms are covered in this article, and as part of the overall strategy, the foundational methods for ransomware defense management and decision-supporting analyses for prevention and mitigation are presented.

Researchers from a wide range of backgrounds, including industrial and academic institutions, have devoted their time and resources to preventing ransomware from accessing the confidential environment. According [5], ransomware is particularly difficult to detect and prevent because of the following issues [6, 7]:

- Ransomware uses encryption as a target and can be easily planted because the system uses many open-source software and services. Injecting malicious files has become very easy.
- Encryption is utilized as a target for ransomware, and it is extremely simple to incorporate into a system because the system makes substantial use of open-source software and services. Malicious files can now be easily injected into a system.
- In order to infect new systems with ransomware, ransomware combines all of the dangerous techniques.
- Most ransomware is seeded via regularly used APIs and is a part of the digitalized ecosystem, making it incredibly powerful and allowing many more infections to penetrate through it.

The goal of this research is to provide a multi-functional technique for detecting ransomware that addresses the aforementioned issues while also using advanced technology applications. Because of the inferring and prognosticating capabilities that AI provides, advanced technologies such as Artificial Intelligence [AI] are being examined. Organizations can become impervious to cyber dangers such as ransomware by utilizing artificial intelligence. Artificial intelligence is widely employed in a wide range of sectors, including as stock market prediction, behavioral analysis, and other applications [8]. Many studies have indicated that artificial intelligence (AI) has the potential to significantly contribute to cyber security. Traditional systems lack the ability to learn from the past in order to recognize data patterns that stimulate ransomware mitigation from the targeted source of infection. As a result, artificial intelligence can be utilized to tackle problems that traditional systems cannot. An erudition-based database of earlier attacks, which collects day-to-day data flow in and out of targeted sources, can be utilized to create data security procedures, allowing the system to detect any malignant behavior.

Section 2 of this paper includes a comparable study that looks at the evolutionary architecture of ransomware, targeted platforms, the impact of a ransomware attack, the importance of data science, and artificial intelligence. The proposed methodology is described in Section 3. Section 4 discusses the

specifics of how the proposed technique would be implemented. Section 5 summarizes the results of the algorithms with specific parameters that were constructed. Sections 6 and 7 provide an analysis of the findings as well as future recommendations.

2 Related Study

The primary goal of researchers and organisations should be to combat escalating cyber dangers and create a secure cyber ecosystem. Cyber-security must be followed in order to secure data amassment tools, methodologies, rule-based policy checks, measures, the security roadmap, threat-mitigation strategies, their respective implementations, recording beneficial practises, and combining new and improved technologies for the betterment of the cyber ecosystem [9]. In recent years, cyber-security research has proved promising and fruitful in terms of implementing preventive measures. Nonetheless, various influential elements must be investigated in order to develop early and improved security threat detection, obviation, and future threat monitoring tactics. In [10], the author describes those constrained susceptibilities that are considered, such as Phishing Attacks (PA), Denial of Service (DoS), and a few malwares; the author also mentions that many vital parameters, such as critical threats, targeted or maltreated application, remediation techniques, and substrate, have not been focused by researchers and organisations. This effort is proposed to focus on the most prevalent vulnerability and take into account the operating environment, targeted gateways, and cloud-based resources or platform weaknesses.

Many cyber-threat detection models have been developed, some of which are still in the early development stages. The author in [11] describes one such breakthrough, in which early threat identification is carried out by an alerting system, informing the administrator to take preventive measures. The goal of deploying cyber-threats is to gain access to sensitive data for the purpose of conducting hostile activities. In [13], the author describes an approach for developing a protected data purposeful model that complies with sensitive data requirements. Modern technologies are not being utilised at the current level of development. As a result, the author proposes for the integration and use of artificial intelligence and block chain technology for threat detection purposes [14]. States that data is the most important asset, and that risks or breaches may be produced by mistakes made by humans. It is possible that humans or data administrators are responsible for the creation of vulnerability

ports as a result of insufficient system architecture and the usability of static operating models.

Cyber-threats are created in order to undermine user trust, inflict harm to organisations, and profit vast quantities of money. To prevent cyber-attacks as soon as they are detected, suitable safeguards for credible data should be put in place [15]. As previously said, cyber-threats are coordinated and designed to gain access to resources for malevolent objectives, and this access is designated by malware [16]. Malware is inserted into specific sources once it has been created (Personal computer, data storage contrivances, third party software, links, cloud, etc.). Despite the fact that there are numerous varieties of malware, ransomware [17] identifies it as the most serious and devastating sort of malware). This paper examines ransomware attacks, which have the potential to cause massive financial or data damage in the cyber world. Ransomware encrypts a user's data in order to exact a hefty ransom in exchange for the contents' recovery. When ransomware infects a database, it jeopardises financial data, business models, and sensitive user data such as personal images and videos [18].

India has been hit by ransomware multiple times; one such red alert instance is mentioned in [18]. Advanced technology utilization has been applied to typical use cases. The red alert was a consequence of impecunious product design, design without security in consideration, low cognizance of the data, and no opportune tracking system. Another such threat is mentioned in [19], which is designated as NEMTY, which infected the windows operating system of the internet users, after gaining access, the NEMTY ransomware encrypted the files, and the ransomware searches for the copy of confidential data, if any and effaces those files as well, leaving no choice to the victim on data recovery and henceforth increase the chances of availing the ransom. This ransomware was query predicated and had its occurrence ecumenical, withal had botnet, but the ransomware specialist used machine learning algorithm to stop the spread of NEMTY ransomware. Another widely spreading and damage-causing ransomware assailment is the WannaCry ransomware, this has evolved over the years, its first occurrence was optically discerned in 1989 [20]. Multiple incipient technologies have been used to fight the WannaCry ransomware but have not been very efficacious [21]. These attacks target the operating systems, entities involved in managing the system, cloud storage, ancillary APIs, and gateways [22]. To mitigate the damage caused by such ransomware. The system that could be held hostage should be designed, considering all the possible parameters and targeted devices.

Established firms, medium scale, small scale businesses, and individuals should pre-plan preventive measures, standardize backups of crucial data, and make progressive efforts to deploy a hard-edge security framework. The paramount data format which have the crucial information, and when the hacker reaches the files with these extensions (.txt, .doc, .rft, .ppt, .chm, .cpp, .asm, .db, .db1, .dbx, .cgi, .dsw, .gzip, .zip, .jpg, .key, .mdb, .pgp .pdf.), the hacker encrypts the files to make it non accessible to its owner [23]. Ransomware can be mitigated, or the damage could be truncated with proper planning. At the same time, product development, taking support of multiple incipient technologies such as high-level tracking, encryption, artificial intelligence, etc., can be beneficial.

In [24], six machine learning classifiers have been used to detect and relegate ransomware. The models can learn and avert attacks, proving that machine learning can be promising to detect and relegate ransomware. Integrating incipient technologies can bring more challenges, such as the lack of ransomware erudition, resources, etc. The hackers who plant the threat are ahead of the organizations because they catch the data activity and understand incipient patterns. Though there is sizably voluminous data to process, these hackers use astronomically immense data processing and develop keenly intellective algorithms to beat the security system [25]. Even with the utilization of artificial intelligence, there is a possibility of breaching the security layer and planting malware. As hackers profit from millions of linked devices, specialized solutions must be built for every entity, application, or program in the cyber world to spot threats from all potential directions [26].

Conspicuously the hackers are accumulating more data to orchestrate better ransomware attacks. One of them is to be able to transmute signatures and become able to cause damage even after receiving the ransom. Huge sectors such as healthcare, financial, and edifying sectors have no cull but to pay the ransom. However, there is an astronomically immense possibility of being chicaned into paying more ransom [27]. After the ransom and data recovery payment, the solution to these unforeseen consequences can be a better backup mechanism, encryption of confidential data, and multilayer checks to avert ransomware from reaching the intended confidential data. Recent advancements in ransomware detection, prevention, and damage control methodologies are to utilize artificial intelligence at every point or gateway that leads to confidential data. In this paper, the utilization of a deep learning algorithm called Long Short-Term Memory [LSTM], which is applied to files with different formats, is proposed. The methodology is constructed considering the environment of operation and the parties involved

in engendering, managing, and forfending sensitive data. Table 1 contains a review on similar research

In this work, multi-type dataset is used, the first type of dataset is used to analyse and detect patterns between multiple malware and ransomware and there is another section is using the patterns from ransomware phishing emails and deploying a multi-data type classifier.

3 Methodology

Ransomware attacks intend to victimize any entity in the digital infrastructure that houses confidential/crucial data. This paper presents a novel methodology to presage and detect ransomware on the user's device, resources deployed by the organization, third-party APIs, and cloud storage. Multiple challenges need to be addressed and solved. The focus of the proposed methodology is layered to surmount the challenges presented in the introduction section. The challenge of the engendering of ransomware utilizing multiple malwares techniques and the mystifying deployment of ransomware as other malware has caused the organization to be tricked into permitting the files into the ecosystem, as these digital ecosystems do not have the facility to incisively capture ransomware coming into the system enmeshed as other malwares. The methodology is split into two components. The first part is dedicated to automatically identifying malware using LSTM that possesses patterns of ransomware. This can benefit the organization to deploy rigorous analysis and support an immediate mitigation plan. The second aspect of this paper demonstrates the application of LSTM to detect ransomware coming in the form of phishing emails. Figure 1 depicts the overall methodology to avail the end-users, organizations, and other entities connected to the world wide web does not fall into malevolent motives set by cyber threats via ransomware attack, which could cause loss of money. Still, it would also decrease the credibility of confidential data. Both the components will have the process working as presented in Figure 1.

Integrating a process is the first step in successfully implementing the working model. The dynamicity of the ransomware occurrence in different formats, be it masked as other threats, malware, phishing emails, or even simple messages, can be hard to detect and track. Consequently, the model is trained on both where malware that has sodality with ransomware and veritable active ransomware infects the resource by a macro that is downloaded along with the attachment that comes with the electronic mail. Both the approaches trick the end-user into cerebrating that the attachment

Table 1 Comparison between the existing works – with data description, results, key takeaway and Our approach

Ref No.	Approach	Data Description	Results	Key takeaways	Our Approach
[28]	Classification of ransomware using Machine learning	Opcodes	5 machine learning algorithms have been deployed with TF-IDF as feature extraction technique. The approach is applied as a binary classifier.	The ML classification accuracy is 99.3%, the work implements a binary classifier.	In this work, deep learning algorithm is used, the dataset used in this work contains opcodes which are extracted from .asm files and is a multi-classifier.
[29]	Ransomware detections using machine learning	Packets collected using PPE Engine	Feature extracted from the packets and used to train ML binary classifier.	Rich-flow records are constantly generated and maintain to increase the features and improvise the accuracy	In this work, LSTM is a deep learning algorithm which learns as the data/packets are feed into the model. Our work adds as an automated process to learn new features and classify ransomware efficiently.
[30]	Windows Ransomware traffic detection using ML	Virus total dynamic files	The features extracted using Tshark is feed into ML algorithms, the results improvise when features extraction is done when compared to directly feeding the data to ML algorithms	Feature extraction plays a critical role, ransomware can me masked into different malwares and hence a patterns exists.	This work attempts to identify the similar patterns between other malwares and ransomware, providing additional features for better classification.
[31]	Framework for detecting ransomware	Malwares from Virus Total	An approach using reverse engineering, statistics and ML with a combined accuracy of 92% is achieved.	Base approach must involve a combination of functionalities which accommodates the existing features, room for new features and a combined framework for optimal classification.	In this work, a combinational approach of having the classifier placed at the entry point of data at every source within the organization with the possibility of learning new data with the deep learning classifier is proposed with the intend to improvise the detection accuracy

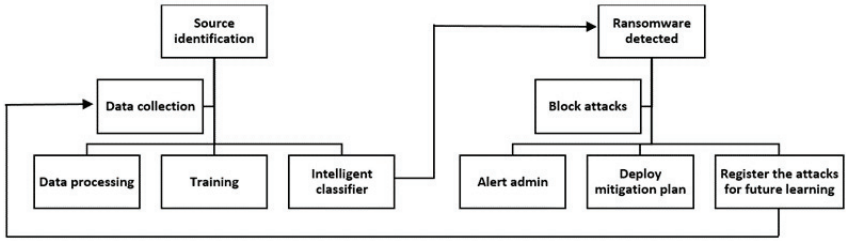

Figure 1 Overall methodology to detect and prevent ransomware within an organization.

or the medium-priority threat alert is not authentic ransomware, that can overtake access of the user's contrivance to the point of controlling all the resources/devices connected, exposing all the devices and infrastructure connected over the network within the organization vulnerably susceptible to threats and data glomming. As presented in the methodology process flow diagram, the algorithm can be deployed on each source and the process to be executed. For example, if the source identified is the PC/laptops utilized by the employees within the organization, the trained algorithm will be deployed on the targeted device, which passed the received files .asm or .mbox (with annexations in word, jpg, js format), identify if there is the occurrence of threats, if yes then the resource will be blocked from the network, an alarm will be raised to the admin. Conclusively, the associated log will be registered and be victualed to the trained deep learning algorithm to improve detection efficiency.

3.1 Source Identification and Data Collection

Ransomware is malware whose intent is to hold resources and ask for ransom. These ransomwares are in high-level cryptic form; collecting such data is possible only through multiple resources. In this paper, the proposed framework will also act as a bottom-up approach with a capturing module set on the target platform on which the intelligent model for ransomware detection is placed. For example, in this paper, the target platform is involved in the organization's internal working, such as cloud services, third party, and internal employees; this is depicted in Figure 2. In this section, the data of different format is considered, and strategies to collect more data for future improvisation is set in place.

Figure 2 has multiple parties that can probably become threats to the organization and act as ransomware blocking and data collectors for improving security against ransomware. Below is the analysis of each of the parties

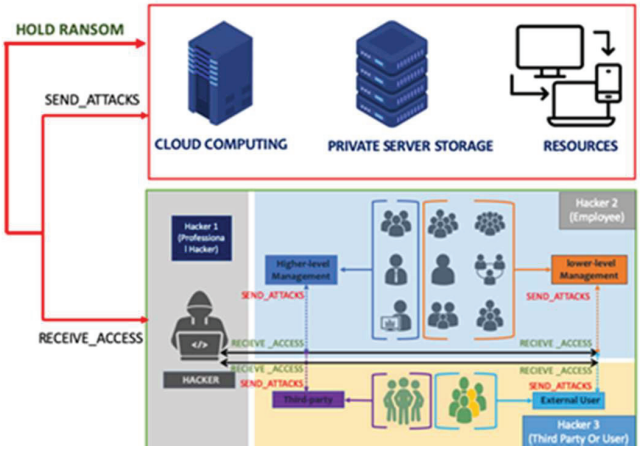

Figure 2 Target platforms that will have intelligent ransomware detection models deployed.

involved in the overall functionality of an organization and the usage of the proposed approach. The risk causers are explained in the following sections:

3.1.1 Third-party services (TPS)

This source has multiple entities such as TPS employees, TPS cloud storage and applications, TPS communication gateways, TPS software and network used to connect to the organization.

Problem: Hackers can send ransomware, making TPS entities a target to reach the organization's sensitive data. TPS can also develop motives to maliciously access the organization's sensitive data by turning into an undercover attack planter, i.e., TPS can be victims or masterminds behind the attacks. Therefore, the TPS can become a threat to the organization in this case.

Proposed solution: To solve this threat, the intelligent detection model is deployed on the enter-point of the organization that the TPS uses. Also, the proposed model deploys a program to collect similar files, which a ransomware expert would later analyze to improvise the security model.

3.1.2 Organization's infrastructure (OI)

Multiple instances have shown that the organization's internal working infrastructure can also contribute to why the organization fell victim to ransom. In this work, the analysis is emphasized on parties involved in the organization's functionality.

Problem: Employees use multiple resources such as Laptops, Virtual workspaces, video conferencing applications, email services and more. All these sources can home multiple unnoticed threats. The ransomware can be masked into different file structures to make the employee open the gate to the organization's most valuable assets. If the employees are not tracked for their activities, they may also be attackers. Since they have more information about the internal working of the organization, the attacks are very hard to negate.

Solution: A robust tracking system and the trained model to detect ransomware in multiple data formats will be deployed on the resources an employee will be working with within the network.

Therefore, all the resources and employees, referred to as entities, are considered as targeted sources for this proposed model. The intelligent model will be deployed to detect the ransomware and inform the organization about the abnormal activities similar to the ransomware patterns. After the sources are identified from the environment, the data is collected in a different format, the most common executable files, i.e., the assembly level files that a user doesn't care about and the phishing emails which have the data.

3.2 Data Processing and Model Training

In this phase, the development is split into two formats (.asm and. mbox format). The data collected from sources are described in the dataset 3.2.1 section and its associated processing. Section 3.2.2 contains algorithms used to create the intelligent detection model.

3.2.1 Dataset

Since ransomware is deployed through different formats mentioned in the literature, in this work, the classification is on the assembly level code format and phishing emails (which contains other formats such as links, .pdfs, and more), there are two sources through which the data is retrieved and processed.

i. Microsoft Big 2015 – Executable data format

Microsoft released this dataset [32], containing 9 malware families and its type. The intent of releasing this dataset was to help data enthusiasts, data scientists, and companies learn about the malicious access deployed on different Microsoft resources.

This work is related to ransomware. Since ransomware is malware, the data mining of which of the nine families could be ransomware was determined and stored as 2 classes, which classifies if the test malware is ransomware or not. The data mining was manually done, and data was prepared according to the problem, this paper is trying to solve. Considering .asm files and the associativity of malwares to be ransomware opens one more perspective to analyze patterns or new data points that an intelligent system must have. Based on the research, Table 2 classifies the malwares associativity as ransomware. Therefore, for this work, the labelled dataset with nine labels is now distributed among 2 labels (Ransomware indexed as 0 class, and other malware indexed as 1 class). This is done for the proposed model to have an idea of the different patterns ransomware can contain, making the proposed model more efficient. Table 2 includes the families belonging to malwares with similar ransomware characteristics in terms of infection, planting, resource attacking, and encrypting files. The characteristics are collected from Microsoft's website as the same company publishes the dataset.

ii. Phishing emails – Multiple formats

The data is collected from [33]; the author has published more than 1800 blogs with active malware and ransomware. The data is processed according to the algorithm in the data processing step, and the training is done for the same. Initially, all the samples are collected in the zip file, then converted the mails into .mbox format to make the algorithm learn effectively and flatten the file by converting it into a txt file. This model can increase its accuracy when more active ransomwares are captured.

The active ransomware collected for training is a mixture of cerber and sage ransomware. The source infects the ransomware after the attachment from the mail is discovered, and an associated executable file runs and downloads the ransomware onto the target source. The working of ransomware infection for the collected active ransomware files is given in Figure 3. As depicted in the figure below, ransomware infection shows the progressive stages of infection that can be detected and blocked when monitored. A pretrained model on the available dataset is deployed on every target device. Malicious file in the form of ransomware reaches the destined target device in the form of an email which will have ransomware available in the form of URL, email header, email body, etc. The attachment, when downloaded in word format or images, parallel auto-downloads the macro which starts the infection process, encrypts the files, and displays the ransom request.

Table 2 Malware families that could be ransomware (Source: Microsoft.com)

Malware Families and Types	Ransomware	Reason
Ramnit (Worm)	Yes	The nature of this malware attack is to steal confidential information, which is one of the ransomware features. Ramnit belongs to the trojan family. Since that family is available in the dataset, we map it to ransomware.
Lollipop (Adware)	No	This Adware doesn't have recorded evidence of data stealing caused by search engines and keywords. Hence there was not much relativity to ransomware.
Kelihos_ver3 (Backdoor)	No	Kelihos is a botnet associated with infecting multiple computer resources, and it's also associated with theft but not holding resources for ransom.
Vundo(Trojan)	Yes	This malware is a trojan and infects a source by sending phishing emails and attachments to reach ransom worthy data.
Simda(Backdoor)	Yes	This malware works by making the machine hostage to carry malicious activity, including holding the resources for ransom.
Tracur (TrojanDownloader)	No	These malwares mainly generate revenue by redirecting to different web sources. No instances of holding resources for ransom were found.
Kelihos_ver1 (Backdoor)	No	This malware is used to extract links and spread the virus to resources. No instance of holding ransomware through these resources was found.
Gatak (Backdoor)	Yes	Multiple instances of Gatak being deployed on the healthcare resources
Obfuscator. ACY (Any kind of obfuscated malware)	No	These malwares have many purposes, such as infecting a computer through spam messages. But no information that helps us link this malware to ransomware was found.

Depending on the criticality of the data to the organization, if the organization decides to pay the ransom, the attacker releases the files. There are chances that the attacker places the same infective macro at a different location in the device to initiate reattacking. This paper proposes that the process will identify such threats at every gateway. With the growing data, the model will gain more and more patterns to accurately block the infection from entering the system. This mechanism can also be repeated after the resources are released to check if there are no additional/hidden ransomware infections

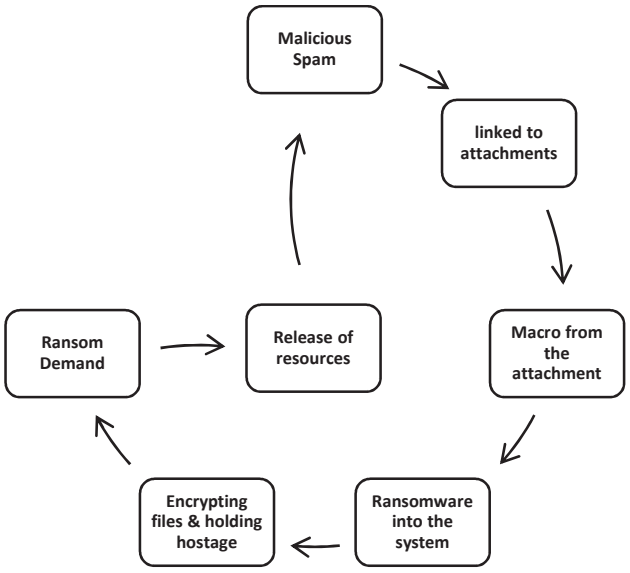

Figure 3 Infection cycle of ransomware deployed via phishing emails.

planted. Therefore, Figure 3 has a feedback loop, which will help the organizations scan for ransomware that could probably be hidden within the hostage resources.

3.2.2 Training AI models

In this work, the deep learning algorithm LSTM is used to demonstrate the working and effectiveness of the system. These models will be the core of the intelligent system.

3.2.3 i. Deep Learning Model 1 – LSTM

Working of LSTM: Dealing with ransomware data collection is a challenging task but can be solved with the help of tools and programs. The more complex challenge that ransomware data imposes is the data processing challenge. The data is of different formats, and multiple patterns will be achieved. The intelligent model will have to keep up with the learning process, therefore, one of the best suitable algorithms for this problem is the Long-Short Term Memory (LSTM) [34]. LSTM, which overcomes the problem of vanishing gradient which occurs during back-propagation, is handled in a series of components: The memory unit, which stores the temporal state of the accomplished network; gates, the input and output gates which

help in modulating the input and output activation and finally the forget gate which helps in resetting the cell's memory as it adapts to the training. The mentioned components constitute to term "LSTM cell." For this work, .asm files are used and the data is preprocessed to have multiple tokens and to map the result given in the experimental result section.

The steps involved are given below, these are used in calculating the mapping between the keywords to tokens and in training the LSTM model to predict the ransomware sequence from the given file. LSTM starts with creating the forget gate, which contains the detected token. The objective of this layer is to forget the old token whenever a new token is detected. RT Represents the mapped tokens to detect the ransomware.

The equation for this is

$$F_{RT} = \sigma(W_x F x_t + W_h F h_{t-1} + W_c F c_{t-1} + B_F) \tag{1}$$

The above formula helps the network forget the previously extracted tokens considering the input and output cells at the time (t–1), i.e., the previously present token in the memory cell.

The input and output gates to learn the ransomware token are modulated using the following Equations (2) and (3), here x_t represents the input at time t. c_t h_t represents the output and current state at time t. $B_F, B_I, B_O,$ B_C represents the respective bias for forget, input, output, and current cells; and W represents the weight of the components.

$$I_{RT} = \sigma(W_{xi} I x_t + W_{hi} I h_{t-1} + W_c I c_{t-1} + B_I) \tag{2}$$

$$O_{RT} = \sigma(W_{xi} O x_t + W_{hi} O h_{t-1} + W_c O c_{t-1} + B_O) \tag{3}$$

$$C_{RT} = \sigma(W_{xi} C x_t + W_{hi} C h_{t-1} + W_c C_{t-1} + B_C) \tag{4}$$

The activation function plays a part in exciting to token that has been trained as a ransomware token. To implement the activation, LSTM has two types of activation functions the Sigmodal activation function used by forget, input and output gates, and the tanh for the current state of the cell [35].

There are many advantages this algorithm has [36], such as the learning rate. It also works well if no prior finite stages are found in the data. Quickly understanding two or more similar pattern occurrences with less time complexity is important for ransomware detection. LSTM can handle massive data. Therefore, this paper proposes using LSTM on Microsoft Big 2015, which contains the files in .asm format. The time complexity of LSTM is calculated using the following equations:

$$Time_{Complexity} = O(UH + UMS + HSF) \tag{5}$$

Where, U = Number of output units, H = Hidden layers, M = Number of memory cells, S = Size of memory cells. The detailed flow chart used for this work is given in Section 3.3.

Optimizing LSTM: Optimizers are put in place to improvise the training of any deep learning algorithm. They hyper tune the training parameters such as weights to reduce the loss/poor performance of the algorithm. Since the data collected to detect ransomware needs parameter tuning and constant weight adjusting, the work is implemented on four optimizers, the resultant of the same is mentioned in the experimental result section.

The selected optimizers are for this work are as follows [37, 38].

- **Adams optimizer:** This is one of the fastest-growing advanced optimizers for the stochastic gradient technique. Adams optimizer can help learn better in datasets that adapt small or large datasets such as the considered in .asm files from Microsoft. The calculative measures for Adams optimizers are as follows:

$$\emptyset_{t+1,i} = \emptyset_{t,1} - \frac{\partial}{\sqrt{\epsilon} + \sum_{r=1}^{t} (\nabla J(\emptyset_{\tau,i})) * 2} \nabla (J(\emptyset_{\tau,i}) \tag{6}$$

- **RMSprop:** This optimizer is relatively associated with Rmsprop, similar to Adam's optimizer, the learning rate is relatively slower than Adam's and has proven to reap the convergence faster. The calculative measure for the same is as follows:

$$\emptyset_{\boldsymbol{t+1,i}} = \emptyset_{\boldsymbol{t,1}} - \frac{\partial}{\sqrt{\epsilon + \boldsymbol{E}[\boldsymbol{g^2}]\boldsymbol{t}}} \boldsymbol{\nabla}(\boldsymbol{J}(\emptyset_{\boldsymbol{t,i}})) \tag{7}$$

- **Adamax:** Adamax is a gradient-based optimizer that adapts to the parameters by adapting its learning rates. It has a more significant update pattern to infrequent data and smaller updates to frequent data.

$$\emptyset_{\boldsymbol{t+1,i}} = \emptyset_{\boldsymbol{t,1}} - \frac{\partial}{\sqrt{\boldsymbol{G_{t,1}} + \epsilon}} \boldsymbol{g_{i,t}} \tag{8}$$

- **Nadam:** Nesterov-accelerated Adaptive Moment Estimation, this optimizer combines Adam and Nag optimizers. This optimizer delays adaption and modifies momentum $m_t = \partial m_{t-1} + \partial \nabla (J(\emptyset_{t,i})$, the equation for the same is:

$$\emptyset_{\boldsymbol{t+1}} = \emptyset_{\boldsymbol{t}} - \boldsymbol{m_t}. \tag{9}$$

In this work, during data preparation, the models learn on a sequence of keywords and tokens mapped during data processing. The deep learning algorithm uses the above optimizer to improvise efficiency. The obtained result is presented in the experimental result section.

3.3 Intelligent Classifier and Ransomware Blocking

The above mentioned deep learning algorithm will be trained on the available data. After these models are trained, they will be deployed on the identified target platform. For example, the intelligent classifier can be deployed on an employee's laptop to block any ransomware containing emails, links, or files. LSTM can be deployed on the cloud to block the hacker from deploying the executable format. These classifiers are deployed on different targeted contrivances, which have been identified in Section 3.1. This approach will allow the organization to increase security and keep confidential data from any targeted platforms and entities out of malicious reach. Ransomware is orchestrated and masked by the cyber-threats. Still, the organizations integrating a keenly intellective detection system at every point of ingression can avert damage to a more preponderant extend and contribute to organizations' safety from these threats and make the data less vulnerable. Still, there are cases where just detection is not adequate. Consequently, a better backup mechanism such as blockchain can be implemented on this proposed system. Due to its immutable comportment, blockchain is the best solution to backup highly confidential data. Blockchain is out of scope from this work. It could be the most efficacious when deployed at an organization whose data leak can make it reach a deadend.

3.4 Data Collection for Future Improvisation

Every research work must be planned, keeping the current situation and the future situation in mind. The challenge faced during data collection can be solved by implementing a data collection strategy to help us achieve better classification. Since data is the source of decision-making, more and more data can give accurate classification. Therefore, a database with the already analyzed ransomware pattern is set on the targeted device. An email is triggered to the admin if there are anomalies detected that the classifier can't solve. The overall methodology is designed to keep the current data collection state, detection techniques, and progressive data collection for better decision-making.

4 Implementation

The implementation will be categorized into two learnings; the algorithm utilized by both the categories is LSTM. The categorization will detect ransomware in both .asm format and .mbox format. This will sanction us to deploy an efficient methodology at the file structure level, as .asm formats, which are assembly-level executable files that are not readable and can reach the operating system of the resource. The other part of the implementation is the methodology at the network level, which receives information via emails, and therefore the resources are prone to fall victim to ransomware attacks.

For category 1 – Machine executable .asm files derived from Microsoft Big Data 2015, this dataset is utilized, as we have analysed and presented in the exordium. The desideratum of finding ransomware is masked in other malware, making it arduous to detect. Since other malware is not damaging to the extent the ransomware does to the organization. Considering the operation priority methods, there is a high chance of the organization de-prioritizing the threat and sanctioning the ransomware to enter the system, as the detection is chicaned into believing that it is malware and requires no immediate shutdown of critical resources which contain the heart (confidential data) of the organization.

After analysing the data presented by Microsoft Big Data 2015, the format these algorithms are alimented with is the .asm format, which is pre-processed and stored into data frames. LSTM overall flowchart is given in Figure 4. A series of data processing, cleaning, and storing in the proper format is performed to abbreviate the resource consumption in terms of storage and time. Since this learning category is focused on finding the malware that could mask the ransomware, the algorithm will associate the malwares possessing the characteristics of ransomware in terms of execution, resource infection, and over-taking control. Predicated on the attributes of 9 families of malware present in the dataset, mapping nine classes into 2 classes, is Ransomware and Other Malware, is done for the model to learn the patterns and variants of malware that can be habituated to plant a ransomware attack.

In LSTM, the positional relationship between words constitutes a sentence, and the context between sentences includes an article. Similarly, the positional relationship between instructions constitutes an assembly function, and the mutual call between functions forms an assembly instruction file. The injunctive authorizations consist of opcodes and operands, which are then victualed to LSTM for learning and have a feedback loop to data for future

Table 3 Mapping malwares into 2 classes

Class	Malware
1	Ransomware
2	Other_Malware

learning. The files are mapped into 2 classes, one class is the Ransomware associated class, which has the patterns of threat sodality. Another class is called Other Malware, which typically is a threat but is not of high priority and can be considered while decision-making on the resource shut down.

Referring to Table 2, column 2 contains the details about, if the malware is ransomware or not, the "Yes" will be mapped to "1" and "No" will be mapped to "2".

The data processing steps are crucial to achieve multi-format ransomware detection, in this paper, the .asm files are read from the source folder iteratively, and initial cleaning is performed to remove unnecessary entries, which would reduce the processing of unnecessary data and increase in efficiency of learning. After extracting the .asm files into text segments, split the lines in the text segment using whitespace characters as the delimiter, and remove unnecessary tokens from the text segment (i.e., anything that is not an opcode or an operand). Since the dataset mainly contains opcodes and operands, the data can be processed to flatten the files to convert line arrays into token arrays. Iterate the keyword and token mapping to reach the best combination, in this paper we use 200,250 and 300 keyword token mapping. The resultant is mentioned in the experimental section. All the malwares from the Microsoft Big 2015 is extracted and mapped to the associated labels mentioned in Tables 1 and 2.

In Figure 2, the .asm files retrieved from Microsoft Big 2015 have certain unwanted data, which is common and when removed, can reduce the processing time, increase the learning rate, and decrease the loss while training. The mapping of the classes from 9, which was retrieved from Microsoft Big 2015 and stored as label.csv is mapped to 2 classes which will be stored as Ransomwarelabel.csv.

LSTM is trained under 4 folds. The optimizer used is rmsprop (selected after training with other optimizers). The number of hidden layers is 32. The activation function used is SoftMax. The average training loss function with the accuracy is generated for validation and presented in the Experimental results section. The results from category 1 focus on helping the organization take advanced steps and corrective actions based on the malware that is

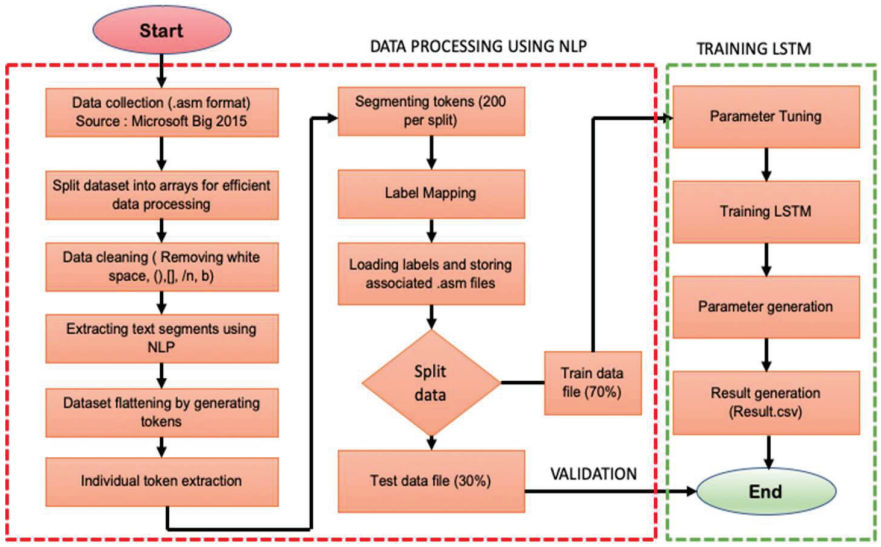

Figure 4 Flow chart of deep learning model – category 1.

mapped to have ransomware patterns and malware that cannot take over the organization's resources.

For category 2, this category learns about pure ransomware collected from the source [34]. The algorithm learns on 578 ransomware emails. These emails contain ransomware in the form of URLs, attachments, and the body of the message. The data is split for the pre-processing, and the lemmatization technique is used to derive the base words from the emails and return the list of base words that form the training data set. For feature selection, we use count vectorise (BOW) and TF-IDF transformer to a set of input features. The data is normalized, flattened, and converted into the array to distribute the train and test data set for the LSTM to learn.

The above figure explains the overall high-level overview of the ransomware learning on the .mbox format received emails. The focus is to analyze each email, attachment or URL that comes via an email before the data is downloaded. With multiple formats, the implementation achieves a resultant model trained on LSTM to detect ransomware enmeshed as malware. This could help the organization detect the ransomware attack, but it can also help plan immediate actions to be taken in time of crisis, as all the resources shut down can cost the organization, the priority-based decision making will help incur the lesser cost.

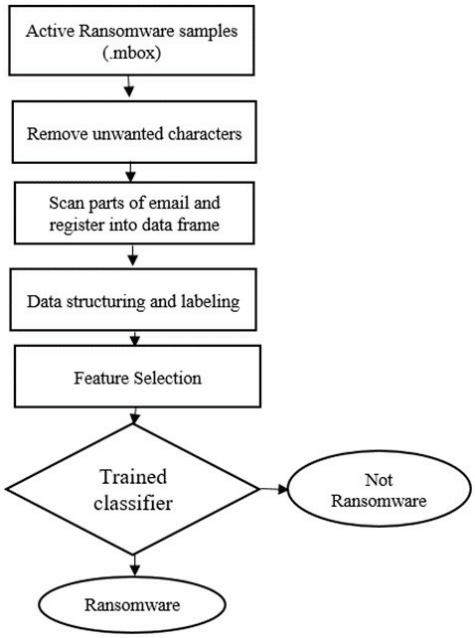

Figure 5 Flow chart intelligent algorithm – category 2.

5 Experimental Result

The paper proposed the below algorithm follow to execute the results. Both ransomware via email and machine readable files are considered in this paper, for the experiment analysis, the test data is passed through the intelligent classifier to detect any patterns that are similar to that of the existing features. Ransomware are hard to detect, hence many features are added to ensure greater accuracy of detection. In this work an extraction of patterns associated to ransomware which is exhibited by other malwares is taken into account along with the ransomware attacks via phishing email. This results in multi-format ransomware detector which can help in placing the right alarming measure. The proposed alarming measures are to notify the admin immediately and to shut down or make the resources with the high-priority flag that identifies the resources which contain the most critical confidential data, this will help in having the other activities function normally and to avoid the reach of ransomware criminal to the destination.

In this section, the model's overall performance on the executable ransomware data collected from Microsoft Big 2015 and Phishing emails

Table 4 High-level algorithm for ransomware detection with multi-format dataset

End2End Ransomware, detection and alarming methodology	
1:	START
2:	Input -> D(x) ▷ # dataset incoming to the source with confidential data in different formats
3:	Dataframe(x) -> Clean(D(x)) ▷ # clean the dataset to remove unwanted characters -> .asm_clean(D(x)) ▷ # cleaning function for dataset with .asm format -> .mbox_clean(D(x)) ▷ # cleaning function for dataset with mail box format
4:	Repeat step3 until the input D(x) is fed into the source.
5:	DF(x) -> Format(D(x)) ▷ # preprocess the data set to be in the desired format -> .asm_format(D(x)) ▷ # formatting dataset function for files with .asm format -> .mbox_format(D(x)) ▷ # formatting dataset function for files with mail box format
6:	Repeat step5 until the formatted input D(X) is feb into the source.
7:	A <- Feature (D(x)) ▷ #extract features using NLP techniques
8:	DLC <- Intelligent_classifier (A) ▷ #Pass the identified features into trained deep learning classifier
9:	Alarm_measure(n) <- Output <- Y_pattern (MLC(A)) ▷ #Execute alarming measures if the pattern is identified
10:	End<-Output <- N_pattern (MLC(A)) ▷ #Execute safe message
11:	END

Table 5 Accuracy comparison

Deep Learning Algorithm	Accuracy
LSTM – Category 1	98%
LSTM – Category 2	87%

collected from the source mentioned in (Malwaretraffic.com), the models' accuracy and loss function have been determined. The result after the training is given in Table 4.

The results obtained by training LSTM on big data were given to the research committee to develop a new and better solution. This experiment aims to recognize the similar patterns of other malwares that can be a ransomware or hold organizations' resources and demand a huge ransom. Figure 7 is the visualization obtained from the deep learning model. To achieve the mentioned accuracy for LSTM, multiple steps were carried out. Given below in Table 6 is the model summary of LSTM for category 1.

The implementation started with the data cleaning process, which structured the data into arrays with only opcodes and operands, which is the expected result at the end of learning, then flattening of each line_arrs, which contains the individual tokens (opcodes and operands), so that text_arr has token_arrs instead of line_arrs. Based on the tokens retrieved, generate a

Table 6 Category 1-dataset training LSTM details

Layer (Type)	Output Shape	Param #
Embedding layer	*32*	*88320*
LSTM	*32*	*8320*
Dense	*2*	*66*
Total params: 96,706	Trainable params: 96,706	Non-trainable params: 0

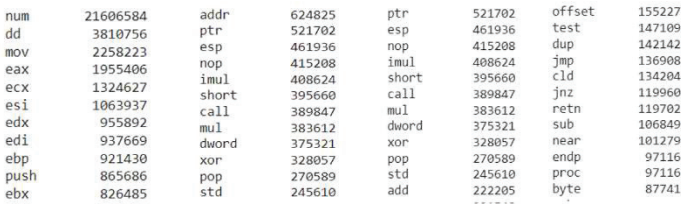

Figure 6 Snapshot of tokens and indexing values.

Table 7 Results after training the LSTM with 200, 250, and 300 token*keyword combination

Token & Keyword Combination	Loss	Accuracy
200 tokens, 200 keywords	18.6	95.76%
200 tokens 250 keywords	22.8	94.04%
200 tokens, 300 keywords	25.1	94.06%
250 tokens, 250 keywords	46.1	86.48%
250 tokens, 300 keywords	54.8	83.89%
300 tokens, 300 keywords	37.6	82.00%

keyword dictionary to map unique tokens found in the dataset to index values. A snapshot of tokens and index values is given below in Figure 6:

The keyword and token mapping of the obtained cleaned data from the .asm files. Proper Token and keyword mapping would increase the learning of ransomware patterns. Given below Table 7 is the obtained accuracy after training the model with the combination of 200,250 and 300 keywords and tokens

The results above show that the algorithm achieves higher accuracy with 200 tokens and 200 keyword combinations. Further, to increase the accuracy and obtain the mentioned accuracy in Table 3. The improvisation is done via optimizers, a set of four optimizers are used to train the algorithm and increase the accuracy while decreasing the loss.

The data collected to detect ransomware needs parameter tuning and constant weight adjusting, the work is implemented on four optimizers; this

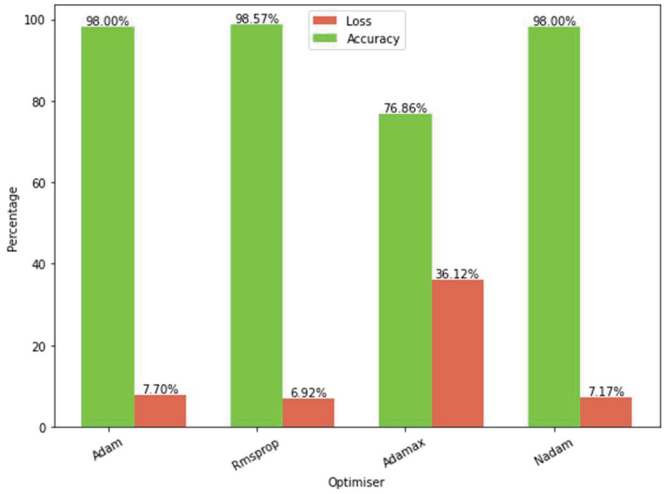

Figure 7 The plot of the optimizers reached accuracy and loss percentage.

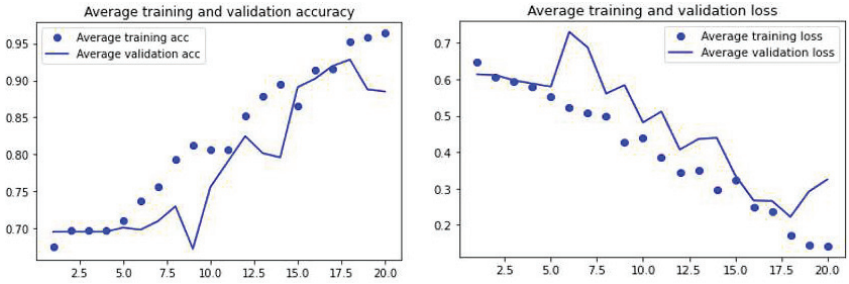

Figure 8 The average training accuracy and associated loss for the ransomware prediction using LSTM.

is done to choose the best optimizer with higher accuracy. After training the algorithm with Adam, Rmsprop, Adamax, and Nadam, choosing Rmsprop to be the best for this work is made. The Figure 7 above mentions the accuracy gained and the loss incurred by each optimizer. The above plot justifies the use of Rmsprop as the selected optimizer for this work. The LSTM is then trained on the dataset to reach the highest accuracy. The plots given below are the achieved results. The validation graph for the model with the 200 keyword-token sequences and rmsprop is shown in Figure 8.

The ransomware prediction on Microsoft 2015 has 98.5% accuracy. The improvement in accuracy is seen after selecting the correct token, keyword combination, and training on the best optimizer.

As mentioned in category 2, the implementation is active ransomwares available in .mbox (email) format. A series of progressive data preparation steps are performed.

1. To collect each word in the email text, the process is to collect the base form of the word and then return the list of base words using lemmatization this has been done.
2. After the list is prepared, feature selection is performed by getting a bag of words for the email text with the function created called bagofwords_transformer.
3. Analysis the transformation is analyzed using the sparsity percentage (Density of dataset population percentage), sparse matrix shape to identify the dimensionality of data, and count of non-zero data points is shown in Table 8.

Table 8 Resultant analysis of the dataset

Sparsity %	(2310, 34237)
Sparse matrix shape	284749
Non-zero numbers	99.9964%

4. Sequencing dataset into an array and preparing a final dataset for training.
5. Training the LSTM and fine-tuning the results. The summary of LSTM training is mentioned in Table 9 below.

Table 9 Category 2-dataset training LSTM details

Layer (Type)	Output Shape	Param #
Embedding layer	*128*	*1280000*
LSTM	*32*	*20608*
Dense	*32*	*1056*
Total params: 1,301,697	Trainable params: 1,301,697	Non-trainable params: 0

LSTM uses “relu” and “sigmoid” activation functions, and the optimizer used is “opt,” after normalization and training on the data point, the achieved resultant accuracy is 86.82%, and the loss is 39.4 %. The dataset has different base text, and therefore this accuracy achieved is shown in Figure 9. The results are expected to improvise as the training with more data is applied.

The outcome of both the machine-executable files and the phishing emails is as expected, the recommendation, according to the methodology proposed,

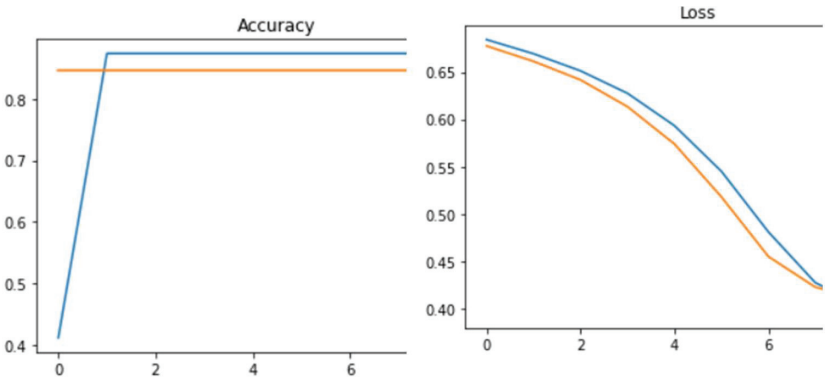

Figure 9 The accuracy and loss from LSTM category 2 training.

is to deploy the intelligent algorithm at every gateway of devices connected to the network and file structure to prevent ransomware, as LSTM is a deep learning model, data can be fed to make it more accurate with passing time and new ransomware email, which the model did not learn.

6 Conclusion

Ransomware attackers plan their attacks in a different format; the traditional ransomware detection methodologies relied majorly on labelled ransomware data. The artificial intelligence models were only trained on the dataset, which is confirmed to be ransomware, but other malicious software could lead to a ransomware attack. This paper proposes a methodology based on the target platforms, which are a crucial part of the organization. The platforms can house multiple vulnerable gaps for the ransomware to be planted and spread the infection. The proposed intelligent classifier is trained on LSTM, which will have the ability to classify 2 different categories of ransomware data placed at two different entry points. The first category is focused on identifying ransomware which tricks the system into thinking that the threat is malware and allows the file into the system with a medium-priority threat. The mapping is done based on the characteristics of the malware family closely resembling the ransomware, the dataset contains machine-readable files, which is the core target of the ransomware criminal, and control of the operating system can be hard to be reverted, which then leads to the organization fall victim to the proposed ransom. This paper proposes to place the category 1 classifier at the file structure level, whenever there

is machine-readable data, the classifier will check for threats and perform actions accordingly, on the other hand, the category 2 classifier is placed at the network layer end to scan the ransomware threats on the received emails, this classifier is trained on live ransomware, the efficiency of detection for category 1 classifier is 98.01%, and category 2 classifier is 86.82%, the efficiency of can be optimized by using more datasets and techniques, which will be a part of enhancement study. The overall framework can be very beneficial to organizations of various sizes.

Future Suggestions

Artificial intelligence has detected various types, and much work on individual formats has been recorded. Multiple formats and pattern matching make the system more intelligent in this work. But there are issues with tracking the devices responsible for allowing the ransomware, the party involved in causing the organization's loss, and yet can't be tracked. There is no backup strategy for confidential data storage. A high-level activity tracking system must be deployed on the intranet and use blockchain to store the files for future enhancement. When organizations face ransom, they can reverse engineering and obtain the confidential data stored on the smart contracts.

References

[1] Corallo, A., Lazoi, M., and Lezzi, M. (2020). Cybersecurity in the context of industry 4.0: A structured classification of critical assets and business impacts. Computers in Industry, 114, 103165. doi:10.1016/j.compind.2019.103165

[2] Z. Liu, K. Choo, W. Liu, and M. Khan, "Guest Editorial: Introduction to the Special Section on Cyber Security Threats and Defense Advance" in IEEE Transactions on Emerging Topics in Computing, vol. 8, no. 02, pp. 264–266, 2020. doi: 10.1109/TETC.2020.2995250

[3] Chesti, I. A., Humayun, M., Sama, N. U., and Jhanjhi, N. (2020). Evolution, Mitigation, and Prevention of Ransomware. 2020 2nd International Conference on Computer and Information Sciences (ICCIS). doi:10.1109/iccis49240.2020.9257708

[4] Homayoun, S., Dehghantanha, A., Ahmadzadeh, M., Hashemi, S., and Khayami, R. (2017). Know Abnormal, Find Evil: Frequent Pattern

Mining for Ransomware Threat Hunting and Intelligence. IEEE Transactions on Emerging Topics in Computing, 1–1. doi:10.1109/tetc.2017.2756908

[5] Fagioli, A. (2019). Zero-day recovery: the key to mitigating the ransomware threat. Computer Fraud & Security, 2019(1), 6–9. doi:10.1016/s1361-3723(19)30006-5

[6] Maigida, A. M., Abdulhamid, S. M., Olalere, M., Alhassan, J. K., Chiroma, H., and Dada, E. G. (2019). Systematic literature review and metadata analysis of ransomware attacks and detection mechanisms. Journal of Reliable Intelligent Environments. doi:10.1007/s40860-019-00080-3

[7] Noor, M., Abbas, H., and Shahid, W. B. (2018). Countering cyber threats for industrial applications: An automated approach for malware evasion detection and analysis. Journal of Network and Computer Applications, 103, 249–261. doi:10.1016/j.jnca.2017.10.004

[8] Humayun, M., Niazi, M., Jhanjhi, N., Alshayeb, M., and Mahmood, S. (2020). Cyber Security Threats and Vulnerabilities: A Systematic Mapping Study. Arabian Journal for Science and Engineering. doi:10.1007/s13369-019-04319-2

[9] Pivarníková, M.; Sokol, P.; Bajtoš, T. Early-Stage Detection of Cyber Attacks. Information 2020, 11, 560. https://doi.org/10.3390/info11120560

[10] Parn, E. A., and Edwards, D. (2019). Cyber threats confronting the digital built environment. Engineering, Construction and Architectural Management. doi:10.1108/ecam-03-2018-0101

[11] Walker-Roberts, Steven, Hammoudeh, Mohammad, Aldabbas, Omar, Aydin, Mehmet and Dehghantanha, Ali. (2020). Threats on the horizon: understanding security threats in the era of cyber-physical systems. The Journal of Supercomputing. 76. 1–22. doi:10.1007/s11227-019-03028-9

[12] S. Yu, G. Wang and W. Zhou, "Modeling malicious activities in cyber space," in IEEE Network, vol. 29, no. 6, pp. 83–87, Nov.-Dec. 2015, doi:10.1109/MNET.2015.7340429

[13] Chikapa, Macdonald and Namanya, Anitta Patience. (2018). Towards a Fast Off-Line Static Malware Analysis Framework. 182–187. doi:10.1109/W-FiCloud.2018.00035

[14] Alenezi, Mohammed, Alabdulrazzaq, Haneen, Alshaher, Abdullah and Alkharang, Mubarak. (2020). Evolution of Malware Threats and Techniques: A Review. International Journal of Communication Networks and Information Security. 12. 326.

[15] Ali, Azad. (2017). Ransomware: A Research and a Personal Case Study of Dealing with this Nasty Malware. Issues in Informing Science and Information Technology. 14. 087–099. doi:10.28945/3707.

[16] Al-rimy Bander Ali Saleh, Maarof, M. A., and Shaid, S. Z. M. (2018). Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions. Computers & Security, 74, 144–166. doi:10.1016/j.cose.2018.01.001

[17] Tailor, Jinal and Patel, Ashish. (2017). A Comprehensive Survey: Ransomware Attacks Prevention, Monitoring and Damage Control. International Journal of Scientific Research. 4.

[18] Simran Sabharwal, Dr. Shilpi Sharma, 2018, Ransomware Attack : India issues Red Alert, International Journal of Engineering Research & Technology (IJERT) Volume 07, Issue 02 (February 2018), http://dx.doi.org/10.17577/IJERTV7IS020074

[19] Bansal, Chetan, Deligiannis, Pantazis, Maddila, Chandra and Rao, Nikitha. (2020). Studying Ransomware Attacks Using Web Search Logs.

[20] Satheesh Kumar, M., Ben-Othman, J., and Srinivasagan, K. G. (2018). An Investigation on Wannacry Ransomware and its Detection. 2018 IEEE Symposium on Computers and Communications (ISCC). doi:10.1109/iscc.2018.8538354

[21] Shakir, Hasan and Jaber, Aws. (2018). A Short Review for Ransomware: Pros and Cons. 401–411. doi:10.1007/978-3-319-69835-9_38

[22] Chen, Q., and Bridges, R. A. (2017). Automated Behavioral Analysis of Malware: A Case Study of WannaCry Ransomware. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). doi:10.1109/icmla.2017.0-119

[23] Luo, Robert and Liao, Qinyu. (2007). Awareness Education as the Key to Ransomware Prevention. Information Systems Security. 16. 195-202. doi:10.1080/10658980701576412

[24] Bae, S. I., Lee, G. B., and Im, E. G. (2019). Ransomware detection using machine learning algorithms. Concurrency and Computation: Practice and Experience, e5422. doi:10.1002/cpe.5422

[25] Bello, I., Chiroma, H., Abdullahi, U. A., Gital, A. Y., Jauro, F., Khan, A., ... Abdulhamid, S. M. (2020). Detecting ransomware attacks using

intelligent algorithms: recent development and next direction from deep learning and big data perspectives. Journal of Ambient Intelligence and Humanized Computing. doi:10.1007/s12652-020-02630-7

[26] Arabo, Abdullahi, Dijoux, Remi, Poulain, Timothee and Chevalier, Gregoire. (2020). Detecting Ransomware Using Process Behavior Analysis. Procedia Computer Science. 168. 289–296. doi:10.1016/j.procs.2020.02.249

[27] Zhang, H., Xiao, X., Mercaldo, F., Ni, S., Martinelli, F., and Sangaiah, A. K. (2019). Classification of ransomware families with machine learning based onN-gram of opcodes. Future Generation Computer Systems, 90, 211–221. doi:10.1016/j.future.2018.07.052

[28] Cusack, G., Michel, O., and Keller, E. (2018, March). Machine learning-based detection of ransomware using SDN. In *Proceedings of the 2018 ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization* (pp. 1–6).

[29] Alhawi, O. M., Baldwin, J., and Dehghantanha, A. (2018). Leveraging machine learning techniques for windows ransomware network traffic detection. In *Cyber threat intelligence* (pp. 93–106). Springer, Cham.

[30] Poudyal, S., Subedi, K. P., and Dasgupta, D. (2018, November). A framework for analyzing ransomware using machine learning. In *2018 IEEE symposium series on computational intelligence (SSCI)* (pp. 1692–1699). IEEE.

[31] Fernando, Damien, Komninos, Nikos and Chen, Thomas. (2020). A Study on the Evolution of Ransomware Detection Using Machine Learning and Deep Learning Techniques. IoT. 1. 551–604. doi:10.3390/iot1020030

[32] Ronen, Royi, Radu, Marian, Feuerstein, Corina, Yom-Tov, Elad and Ahmadi, Mansour. (2018). Microsoft Malware Classification Challenge. doi:10.13140/RG.2.2.34695.91045

[33] Ait Hammou, Badr; Ait Lahcen, Ayoub; Mouline, Salma (2020). Towards a real-time processing framework based on improved distributed recurrent neural network variants with fastText for social big data analytics. Information Processing & Management, 57(1), 102122–. doi:10.1016/j.ipm.2019.102122

[34] Hochreiter, Sepp and Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735–1780. doi:10.1162/neco.1997.9.8.1735

[35] Li, J., Mohamed, A., Zweig, G., and Gong, Y. (2015). LSTM time and frequency recurrence for automatic speech recognition. 2015

IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). doi:10.1109/asru.2015.7404793
[36] Kingma, Diederik and Ba, Jimmy. (2014). Adam: A Method for Stochastic Optimization. International Conference on Learning Representations.
[37] Ruder, Sebastian. (2016). An overview of gradient descent optimization algorithms.
[38] https://www.malware-traffic-analysis.net/
[39] Zhang, H., Xiao, X., Mercaldo, F., Ni, S., Martinelli, F., and Sangaiah, A. K. (2019). Classification of ransomware families with machine learning based onN-gram of opcodes. *Future Generation Computer Systems*, *90*, 211–221.

Biographies

B. N. Chaithanya received the bachelor's degree in computer engineering from Visesvaraya technological University in 2008, the master's degree in computer networks from Visesvaraya technological in 2014 and pursuing PhD in Computer Science Engineering from GITAM University, respectively. Currently working as Assistant Professor in Department of Computer Science and Engineering at GITAM School of Technology, Bangalore. Areas of Interest are Network Security, Threat intelligence, Robotic process Automation and Cyber Security. Published papers in the those specified areas.

S. H. Brahmananda received the bachelor's degree in computer engineering from Sri Siddhartha Institute of technology, VTU University in 1995, the master's degree in computer engineering from NITK, Suratkal in 2004, and the philosophy of doctorate degree in Computer Science Engineering from Dr MGR University in 2013, respectively. He is currently working as an Professor at the Department of Computer Engineering, GITAM University. His research areas include Cyber Security, deep learning, Threat intelligence and social network analysis. He has been serving as a reviewer for many highly respected journals.