International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 08 | August - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Delay Aware Recurrent-Convolutional Neural Network (R-CNN) for Ransomware Detection

[1]Tukur Adamu Muhammad, [2]Muhammad Lamir Isah, [3]Danlami Mohammed, [4]Atiku Baba Shidawa, [5]Mustapha Lawal Abdulrahman, [6]Goteng Kuwunidi Job, [7]Ismail Zaharaddeen Yakubu

[1,2,3,4,5,6] Department of Computer Science, Abubakar Tatari Ali Polytechnic, Bauchi
[7]SRM Institute of Science and Technology Kokalta, India

Abstract - The adverse impact caused by ransomware on computing systems poses a major threat to everyday users and society in general. With continuous growth in ransomware, and newer malicious families emerging every month, the need for strong defensive methods increases every day. While expert-based systems are developed over time, this rate of growth in ransomware creates a need for self-evolving methods of defence that can learn from available data and improve over time. Deep learning methods, in particular, can provide this ability to improve learning with the increasing availability of data. This research proposes an alternative ransomware detection technique using deep learning, specifically Recurrent-Convolutional Neural Network (R-CNN), which will speed up the training time and extract high level features to give more accurate classification. Consequently, our model has improved the classification accuracy of the existing systems to 98.00%. The comparisons with other state-of-the art peer approaches have proven that our empirical model is promising.

Key Words: Ransomware, R-CNN, Deep learning, Neural Network

1.INTRODUCTION

Ransomware has become a significant global threat with the ransomware as-a-service model enabling easy availability and deployment, and the potential for high revenues creating a viable criminal business model. Individuals, private companies or public service providers e.g., healthcare or utilities companies can all become victims of ransomware attacks and consequently suffer severe disruption and financial loss (Alhawi, 2018).

Ransomware is a group of malwares which aims to make computing resources unavailable to the user and demands a ransom amount to make it available again. It has been around for the last decade, but the integration of computers into all aspects of daily life of humans have made them more profitable and hence there is an increasing number of attacks seen against organizations as well as general computer users (Anjana, 2017).

Ransomware can be considered as a serious threat when it comes to protection of information assets. The main targets are internet users. Ransomware hijacks user files, causes difficulties and then requests some funds through extortion for decryption purposes (Bhattacharya & Kumar, 2017). Ransomware can be categorized as malware which can affect the vulnerability of the user's system, allowing the system to be accessible individually and eventually encrypts all the files that have been targeted (Muslim, *et al*., 2019). In 2021, the average cost of recovery and ransom associated with a ransomware attack has been two times more than the 2020 average global ransom demand. During the first two fiscal quarters of 2021, not only did ransomware attacks continue to become more targeted and sophisticated, but the most prolific "Double Extortion" ransomware operators have been observed holding enterprise networks hostage for eight figure sums of up to $40M USD (Herjavec Group, 2021).

Securing our digital assets has become increasingly challenging as our reliance on rapidly evolving technologies continues to grow. The security perimeter in computing has changed from a well-defined boundary that was relatively easy to identify and defend, to an elastic boundary that helps the threats to evolve constantly. Cloud computing provides virtual resources to its consumers through internet. Security in cloud computing is still a hurdle. Securing data, examining the utilization of cloud by the cloud computing vendors, etc are the security issues surrounded by cloud computing. The wide acceptance of the world-wide-web (www) has raised security risks along with the uncountable benefits, so is the case with cloud computing (Anjana, 2017).

Ransomware is known as the most popular Cybercrime in the world (Krunal, 2017). Generally speaking, ransomware is a category of malware that spreads like a worm and inhibits or limits users from accessing their system either by locking the systems screen or encrypting and locking users' files unless after ransom is paid. There are several types of ransomwares and they have been categorized into three basic types (Yakoob *et al*., 2017).

There are five phases of ransomware attack (Quinkert, *et al*, 2018) are described in the Figure 1 below.

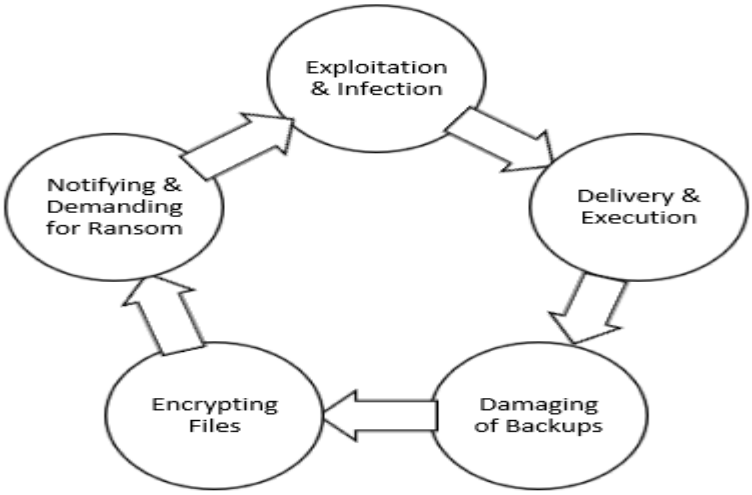

Fig. 1*: Phases of Ransomware Attack*

The first ransomware phase is exploitation and infection. This exploitation is executed through exploit kit and phishing email. Its distribution spread through phishing schemes involving email attachments or downloads. The second ransomware phase is delivery and execution. This is the phase where the ransomware executing files arrive in the computer systems of the victims (Zhanhui et al., 2017) and start the attacking process. This phase only takes a couple of minutes to complete. This process will encrypt key servers in order to retrieve the loss data. Third ransomware phase is damaging of backup files. The ransomware will search for f for important files in the system such as JPG, Doc, and PDF. It will also seek and damage folders including the hidden ones where the backup files are stored (Zhanhui *et al*., 2017).

The purpose of damaging the files is to prevent computer users from performing backup restoration. The fourth ransomware phase is encrypting of files. Criminals will move and rename the target files. After that, they will encrypt and rename the files after a successful encryption (Kok, *et al.*, 2017). When the backup files cannot be opened by users, criminals will perform secure key exchange. This key will give command to users and control the server. The last ransomware phase is notifying users and demanding for money. Criminals will notify users about the payment the latter have to make after the ransomware has deleted the backup files. The demand for payment will be prompted with the payment instructions to clear the ransomware (Zhanhui *et al.*, 2017). After locking the files or system, criminals will demand for payment which is usually high. Usually, the ransom value to unlock the infected files will be raised if the payment is not made within the stipulated time.

Bitcoins are used by ransom operators for payment options such as iTunes and Amazon gift cards Ransomware started to really take off by combining capabilities such as more powerful asymmetric encryption methods and using the new cyber currency of Bitcoin as payment. Satoshi Nakamoto invented the digital asset and payment system, bitcoin and released as open-source software in 2009 (Anjana, 2017).

2. RELATED WORK

The adverse impact caused by ransomware on computing systems poses a major threat to everyday users and society in general. With continuous growth in ransomware, and newer malicious families emerging every month, the need for strong defensive methods increases every day. While expert-based systems are developed over time, this rate of growth in ransomware creates a need for self-evolving methods of defense that can learn from available data and improve over time. Deep learning methods, in particular, can provide this ability to improve learning with the increasing availability of data (Agrawal *et al*, 2019).

Deep Learning algorithms have shown unprecedented success in various domains such as image classification, natural language processing, and speech recognition (LeCun *et al.*, 2015). Following this trend, Deep Learning algorithms have also been applied to malware detection and classification tasks using static and dynamic analysis data exploiting its temporal, spatial, or spatio-temporal structure (Oliveira & Sassi, 2019). It is a subfield of machine learning which employs algorithms trying to imitate the functioning human brain through multiple layers of neural networks. As a matter of fact, recent years have brought a significant development in the area of neural networks, mostly under a paradigm of deep learning. This paradigm encompasses a movement towards creating neural networks with a high number of layers to model complex functions of input data (Kolosnjaji, *et al*, 2016)

A neural network is an interconnected assembly of simple processing elements, *units* or *nodes*, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-unit connection strengths, or *weights*, obtained by a process of adaptation to, or *learning* from, a set of training patterns (Gurney, 2004).

Recurrent Neural Networks (RNN) are a type of neural networks designed for identifying patterns in the sequence of data. In addition to forwarding the output to the next layer, it may feedback as input with the next input vector. These recurrent connections add a state to the network and allow the network to learn broader abstractions from the input sequence. LSTM network is an RNN that overcomes the vanishing gradient problem for large input sequences. A process makes a very large number of API calls and for this large sequence length, LSTM networks are best suited (Maniath, 2017).

The remaining parts of this paper includes the proposed model which is found in section II while section III presents the results and finally section IV is the conclusion.

2. Methodology

Contrary to the convolutional layer, Recurrent Neural Network (RNN) is able to capture long-term dependencies even with one single layer (Hochreiter, 1997). The Recurrent layer has the ability to remember important information across long stretches of time (Hassan & Mahmood, 2017). This research intends to exploit the CNN and RNN in a single architecture to overcome the problems of long-term dependencies in the conventional CNN based architectures. RNN has the potential to memorize long-term dependencies, and therefore able to complement the feature extraction ability of CNN when used in a layered order. LSTMs have the capacity to selectively remember patterns for a long duration of time and CNNs are able to extract the important features out of it. RNN-CNN layered structure, when used for classification, has an edge over conventional CNN classifier (Nagda, *et al*, 2019).

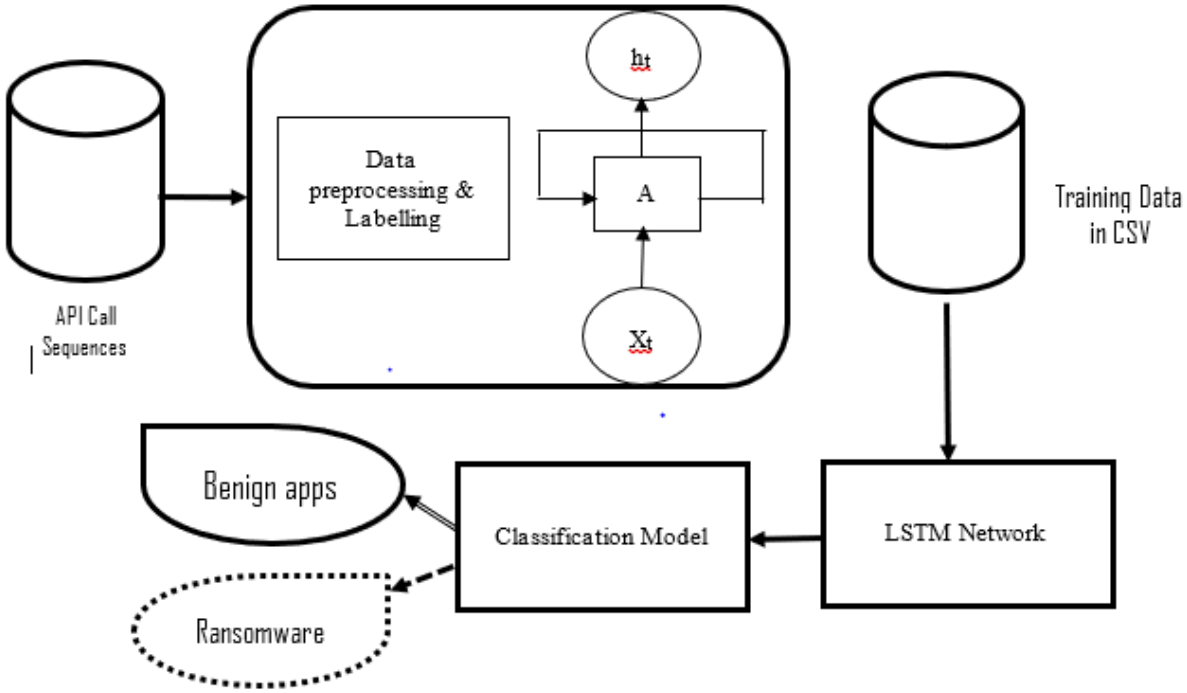

Fig. 2: *Architecture of the Proposed system*

A. Algorithm of the proposed system

As illustrated in Figure 2 above, the proposed model has eight sequential steps from data gathering to detection.

Step 1: Malware samples and Benign (Goodware) files are fed to a Cuckoo Sandbox environment

Step 2: The Cuckoo, in turn, runs the PE files and generates raw JSON reports such as API call sequences,

Step 3: The API call sequences are then fed into CNN-RNN Model for preprocessing, labelling and high level feature extraction.

Step 4: The post-processed data are identified and converted into ordinal categorical values, or comma separated values (CSV)

Step 5: These form the final input that can be used to train the LSTM network.

Step 6: The LSTM network will accept the single CSV file and choose the best model based on accuracy during the training stage.

Step 7: The classification model chosen by LSTM will then classify sequences in the testing stage

Step 8: Finally, binary classification of the sequences as either benign or ransomware is done

B. Description of the proposed system

In order to maximize the utilization of the possibilities given by neural network methodology, we combine convolutional and recurrent layers in one neural network. Figure 2 depicts our neural network architecture. The convolutional part consists of convolution and a pooling layer. The convolutional layer serves for feature extraction and captures the correlation between neighboring input vectors and produces new features. Outputs of the convolutional part of our neural network are connected to the recurrent part, which will be used to explicitly model the sequential dependencies in the kernel API calls traces, thereby extracting features of highest importance from output and reducing the complexity of further data processing.

Convolutional layer learns to extract higher level features that are invariant to local translation, the network can efficiently extract high level features from input sequence. However, it requires stacking multiple convolution layers in order to capture long-term dependencies. Recurrent layers are expected to preserve ordering information even with one single layer. As a result of this observations, the proposed model combines the convolutional and recurrent layers into one single method to efficiently capture long-term dependencies in the input, to perform classification tasks, and to use a recurrent layer as a substitute for pooling layer in the convolutional network to hypothetically reduce the loss of details in local information, capture long-term dependencies and shorten the

processing time. It is expected that this extra feature (integration of RNN into CNN) will give the model capabilities to speed up processing time and obtain high-level feature extractions to capture long-term dependencies in sequence of input dataset.

C. Experimental Setup

The proposed model is implemented in MATLAB R2021a on a machine with the following configurations: Intel Core (TM) 2 Quad CPU 2.33GHz RAM 2GB, 32-bit operating system. The partitioning of data for modelling has no specific ratio, thus initial experimentation is used to test several ratios including 70% for training and 30% for testing the efficiency of the build model.

3. RESULTS

This section discusses the training and performance evaluation of the proposed model.

D. Training Process

The dataset was split into training, testing and validation. 70% of data was set for training, 15% for testing and 15% for validation. To build the network, we transpose the input from columns to rows. An unsupervised training method of the deep neural network (integration of RNN into CNN) was used, the CNN was trained with hidden layer (neurons) of size 10. We set the L2 weight regularization to 0.001, sparsity regularizer to 4 and sparsity proportion to 0.05. The parameter (epoch) used is 7 of 7 with 56 iterations per epoch and completed the total number of iterations 392, we estimate the ransomware types using the deep network. Finally, the overall percentage of correct and incorrect classification was obtained.

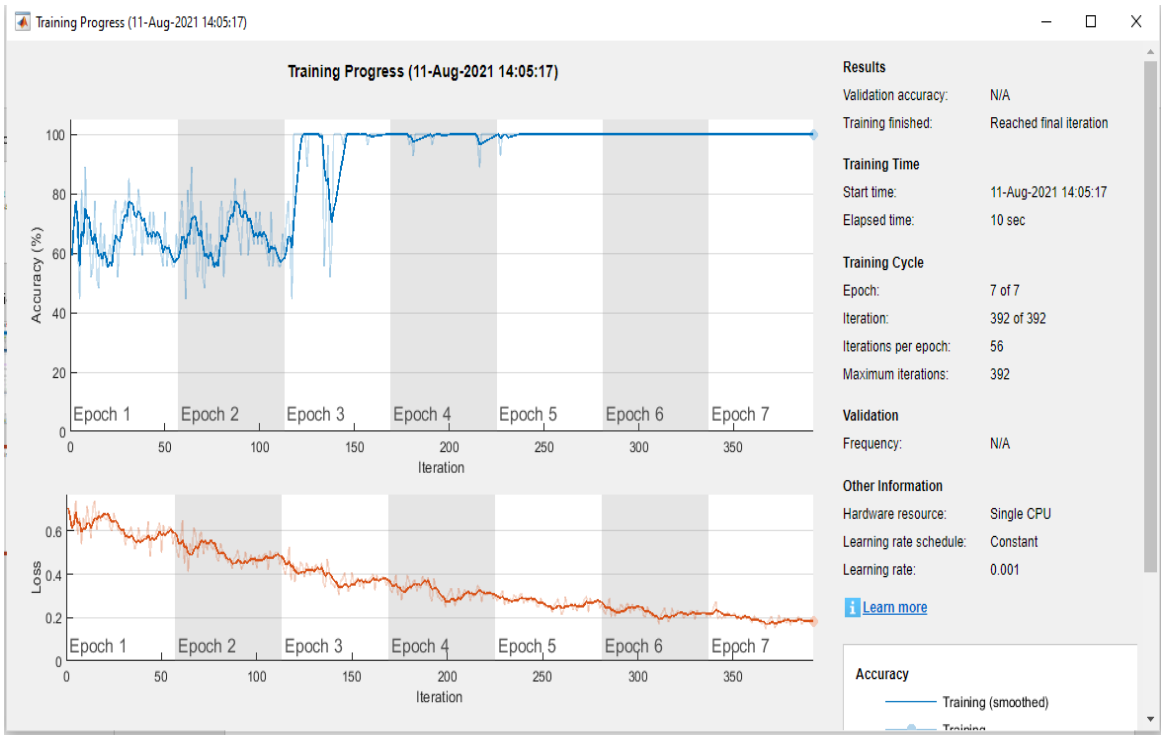

Fig. 3: Training Process graph

E. Performance Evaluation

The performance metrics of the evaluated deep learning approach is shown in table 1 below. The computed metrics shows best results. With this learning model, the proposed system attains a detection rate of (True Positive Rate) of 98.0% with a near zero false positives. In this study we are considering three performance metrics including accuracy, sensitivity and specificity.

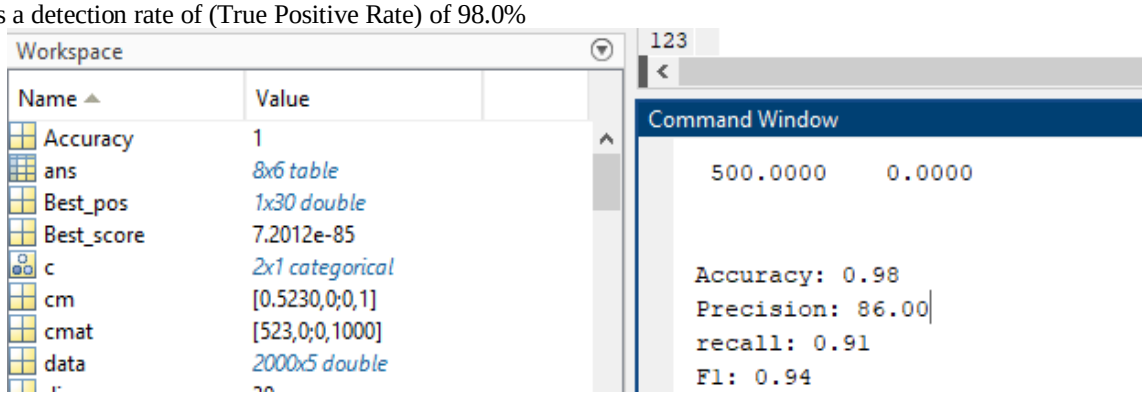

Fig. 4: Performance results

The result in Figure 4 shows an accuracy of 0.98, precision of 86.00 with a recall of 0.91 and F1 measure of 0.94.

To show the performance of our classification system compared with state-of-art existing systems, we investigated similar approaches that have been previously proposed. From the deep learning–based methods to the general classification–based methods, various kinds of the Windows malware detection methods were surveyed. Table 1 shows the results of the investigations. Many existing methods utilize the dynamic analysis method and conduct on various deep learning algorithms. As shown in Table 3, the classification accuracy of our proposed method is higher than the other methods and this study is unique in multiclass classification.

Tab

le 1: Accuracy rate of the Proposed R-CNN Model against Existing Models

Models	**Accuracy Rate (%)**
Alhawi 2018 (J48)	97.10
Bae 2019 (KNN)	96.13
Jogin 2018 (CNN)	85.97
Maniath, 2017 (CNN)	96.67
Proposed System (R-CNN)	**98.00**

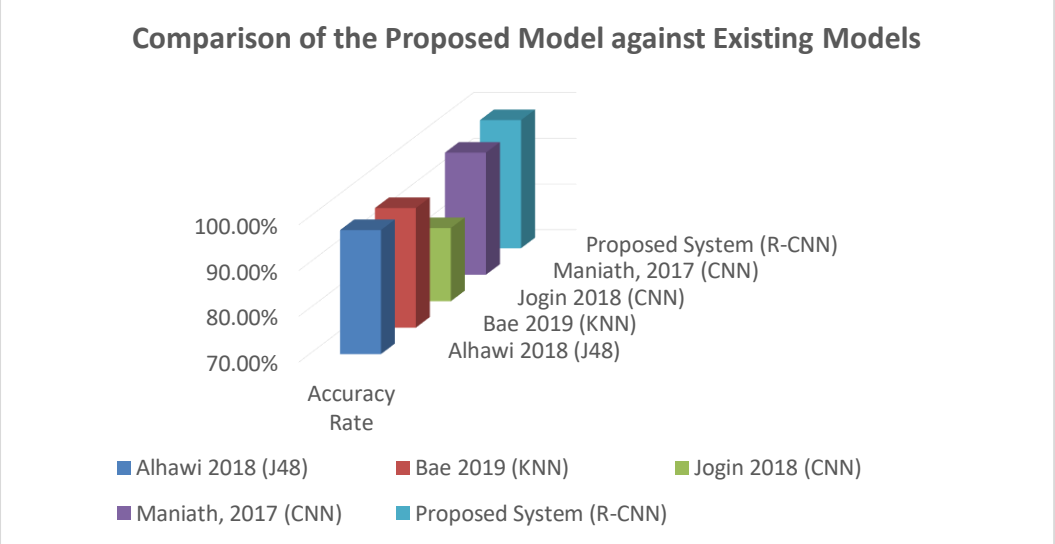

Fig.5 *Accuracy of Proposed System against Other Systems*

Our proposed model achieves comparable high accuracy. We achieve better results compared to the existing models as illustrated in Table 1. The proposed model uses recurrent layers as substitutes to pooling layers, which help maintain the detailed information and perform better. The proposed model is more compact due to the small number of parameters, and less disposed to over-fitting. Hence, it generates better when the training size is limited.

4. CONCLUSION

In this research we showed that, despite the impressive performance of deep learning techniques (such as CNN, RNN, etc) in ransomware detection and classification, many layers of CNN cause training take longer time. Contrary to the convolutional layer, Recurrent Neural Network (RNN) is able to capture long-term dependencies even with one single layer, and has the ability to remember important information across long stretches of time. Therefore, we integrated the CNN and RNN in a single architecture to speed up the training process and overcome the problems of long-term dependencies in the conventional CNN based architectures.

Consequently, our model has improved the classification accuracy of the existing system from 96.67% to 98.00%. The comparisons with other state-of-the art peer approaches have proven that our empirical model is promising.

Reference

Agrawal, R., Stokes, J. W., Selvaraj, K., & Marinescu, M. (2019). *Attention in recurrent neural networks for ransomware detection.* Paper presented at the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Ali, A. (2017). Ransomware: A Research and a Personal Case Study of Dealing with This Nasty Malware. Issues in Information Science & Information Technology, 14, 87–99. https:// doi.org /10.1080/ 13880290 490480167

Anjana, T. (2017). Discussion On Ransomware, Wannacry Ransomware and Cloud Storage Services Against Ransom Malware Attacks. *International Journal for Research Trends and Innovation, 2*, 310-314.

Bhattacharya, S., and C. R.S. Kumar. 2017. "Ransomware: The CryptoVirus Subverting Cloud Security." 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies, ICAMMAET 2017 2017–Janua: 1–6.

Hassan, A., & Mahmood, A. (2017). *Efficient deep learning model for text classification based onrecurrent and convolutional layers.* Paper presented at the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA).

Hampton, N., Baig, Z., &Zeadally, S. (2018). Ransomware behavioural analysis on windows platforms. Journal of information security and applications, 40, 44-51.

Harnedy, R. (2016). 3 better ways to use backup to recover from ransomeware. Retrieved from Barkly: https://blog.barkly.com/3-better-ways-to-use-backup-to-recover-from-ransomware

Kok SH, Azween A, Jhanjhi, & Mahdevan S (2019) Ransomware, Threat and Deyection: A Review. *International Journal of Computer Science and Network Security. Vol.19. No. 2*

Kolosnjaji, B., Zarras, A., Webster, G., & Eckert, C. ((2016, December)). Deep learning for classification of malware system call sequences. *Australasian Joint Conference on Artificial Intelligence*, 137-149.

Krunal, G. (2017). Survey on Ransomware: A New Era of Cyber Attack. International Journal of ComputerApplications,68(3),975–8887. Retrieved from https://pdfs.semanticscholar.org/71df/288033380d3023f09d49b7b55a77677d27a2.pdf

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Maniath, S., Ashok, A., Poornachandran, P., Sujadevi, V., Sankar, A. P., & Jan, S. (2017). *Deep learning LSTM based ransomware detection.* Paper presented at the 2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE).

Muslim, A. K., Dzulkifli, D. Z. M., Nadhim, M. H., & Abdellah, R. H. (2019). A Study of Ransomware Attacks: Evolution and Prevention. *Journal of Social Transformation and Regional Development*, *1*(1), 18-25.

Nassi, B., Shamir, A., &Elovici, Y. (2017). Oops!... I think I scanned a malware. arXiv preprint arXiv:1703.07751.

N. Khoa, T. Dat, M. Wanli, S. Dharmendra, \An approach to detect network attacks applied for network forensics," in 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'14),pp. 655{660, 2014.

Oliveira, A., & Sassi, R. J. (2019). Behavioral Malware Detection Using Deep Graph Convolutional Neural Networks. In.

Yaqoob, I., Ahmed, E., Hashem, I. A. T., Ahmed, A. I. A., Gani, A., Imran, M., &Guizani, M. (2017). Internet of things architecture: Recent advances, taxonomy, requirements, and open challenges. IEEE wireless communications, 24(3), 10-16.

Zhanhui, L., Azlina, N., & Rahman, A. (2017). A Review on Ransomware Trend of Attacks and Prevention