iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.45172

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Using Neural Network Techniques to Detect Malicious URL for Preventing Ransomwares

Sonal Yadav[1], Neelam Sharma[2], LalitKumar P. Bhaiya[3], Virendra Kumar Swarnkar[4]
[1]*M.Tech.CSE Scholar, BCET, Durg*
[2]*Assistant Professor, Department of CSE, BCET, Durg*
[3, 4]*Associate Professor, Department of CSE, BCET, Durg*

Abstract: ***Ransom ware is a kind of malignant malware programming that takes steps to distribute or hinders admittance to information or a PC framework, for the most part by scrambling it, until the casualty pays a payoff expense to the assailant. As a rule, the payoff request accompanies a cutoff time. Presently days and assailants executed new strategies for effective working of assault. The World Wide Web has become an important part of our everyday life for information communication and knowledge dissemination. Malicious URLs host unsolicited content (spam, phishing, drive-by exploits, etc.) and lure unsuspecting users to become victims of scams (monetary loss, theft of private information, and malware installation), and cause losses of billions of dollars every year. To detect such crimes systems should be fast and precise with the ability to detect new malicious content. The current work focuses on developing a model using convolutional neural networks for efficiently preventing ransom wares.***
Keywords: Ransomware, Ransomware Detection Techniques, Convolutional Neural Networks (CNN).

I. INTRODUCTION

Ransom ware assaults are assault that scrambles or locks your documents or frameworks with the assistance of one of the cryptographic calculation like AES, RSA and request that the client pay a payoff to get back your records or framework in working. The assault is exceptionally famous and having one of the most assaulted lately on network security. Ransom ware attack identification frameworks are exceptionally famous and extremely valuable in the Attack Countermeasure methods in the organization security. A few instances of Ransom ware incorporate [6]:

1) *WannaCry* – A strong Microsoft exploit was utilized to make an overall ransomware worm that tainted north of 250,000 frameworks before a killswitch was stumbled to stop its spread.
2) *CryptoLocker* – This was one of the first of the current age of ransomware that necessary digital money for installment (Bitcoin) and scrambled a client's hard drive and joined organization drives. Cryptolocker was spread by means of an email with a connection that professed to be FedEx and UPS following warnings.
3) *NotPetya* – Considered one of the most harming ransomware assaults, NotPetya utilized strategies from its namesake, Petya, for example, tainting and scrambling the expert boot record of a Microsoft Windows-based framework. NotPetya utilized a similar weakness from WannaCry to spread quickly, requesting installment in bitcoin to fix the changes. It has been characterized by some as a wiper, since NotPetya can't fix its progressions to the expert boot record and delivers the objective framework unrecoverable [4].

The major approaches for detecting and preventing ransom wares are as given below:

A. Local Static Information

A location calculation dependent on nearby static boundaries is fit for recognizing malware before it runs. A powerful calculation dependent on nearby static boundaries is the best and keeps away from any deficiency of client information. The static data got from the documents is connected with text strings or capacity calls.

1) *Text Strings:* Common strings found in ransomware doubles are "recover", "bitcoin", or "encode". It can likewise contain notable space names or IP addresses. The counter malware programming can look for catchphrases or set expressions.
2) *Function Calls:* The most widely recognized capacity calls found in ransom ware programs are connected with cryptography calculations and document access. They can be capacities from notable powerful framework libraries or statically connected libraries.

B. Local Dynamic Information

Dynamic data can be measurable in nature; thusly, it requires gathering tests of ransom ware activities during a specific timespan. Disregarding genuine malware (bogus negative) and hindering harmless programming (bogus positive) should be considered as lethal calculation blunders. It tends to be assembled into three classes [9]:

1) *Information Access Data:* These boundaries are connected with the alteration of the substance in client records.
2) *Metadata Access Data:* They measure the activities taken by the ransomware on client records, not the substance of the documents, but rather how and when the documents were altered.
3) *Function Calls:* They measure the real library or framework capacities called by the presume cycle.

C. Information Extracted from Network traffic

Network traffic can be acquired at a contaminated host or at the nearby organization Internet access interface. Against malware programming can break down this traffic and recognize ransom ware activity. On the off chance that the activity is past to the information encryption stage, it can hinder the ransomware before it makes disastrous moves.

II. RELATED WORK

An early detection structure for detecting crypto-ransomware families has been proposed [12]. The structure includes 3 modules, especially pre-processing, including design and identification modules. The pre-processing module uses a sliding window convention, FCM calculations are used for salience extraction, and anomaly-based findings are used for structure formation. [13] Proposed a framework that uses sample sequential matching computations for optimal element selection to orchestrate ransomware from benign applications. Using Locky (517 samples), Cerber (535 samples) and TeslaCrypt (572 samples) three kinds of ransomware tests, the detection accuracy on Goodware is close to 100%, and the accuracy in localization reaches 96.5%. The detection time of less than 10 seconds was tested using J48, MLP, Sacking and Random Backwoods grouping calculations. Another approach is to use AI methods combining static and dynamic investigations to identify ransomware, namely RansHunt [34]. The framework separates the most relevant highlights from ransomware and ranks ransomware using Goodware's Support Vector Ransomware Machine Calculations. Additionally, 93.5%, 96.1%, and 97.10% of the 1,283 ransomware, goodware, and scareware samples, respectively, implemented static, dynamic, and hybrid methods to detect ransomware attacks.[25] Use the Malware-O-Matic investigative phase to create an arrangement that validates the actions of the document framework, i.e. data-aware defenses, to gather ongoing information. They attempted measurable ransomware identification tests on more than 798 ransomware family samples and reported 99.37% accuracy, with each test losing up to 70MB 90% of the time and reaching 7MB 70% of the time. Testing was performed with PNG, ZIP and PDF documents. [37]. Introducing REDEMPTION, an original protection method that creates a working framework to quickly recover from ransomware attacks. The framework examines all I/O usage requests under each process premise for possible signs of ransomware attacks. Assuming the I/O request shows signs of a ransomware attack, the I/O request will be terminated. This allows for zero information incidents. The dataset used for the localization process consisted of ransomware tests and benign samples, 504 samples from 12 dynamic ransomware families were used as ransomware families, and more than 230 GB of innocuous information was collected.

III.PROBLEM IDENTIFICATION

Signature-based, behavior-based, and specification-based techniques are commonly used to detect malware. Signature-based detection techniques maintain a database of known malicious program signatures and use those signatures to identify the presence of an attack by matching signatures stored in the database. This technique provides fast malware detection and requires less computing resources. However, it cannot detect new or unknown malware. Behaviour-based detection techniques analyse various characteristics, such as the source and destination of malware, attachment types, and statistical characteristics. Behaviour-based methods can detect both known and unknown malware, but they require significant computing resources, such as memory and CPU time.

IV.METHODOLOGY

Attackers additionally use emails and social media systems to distribute ransom ware with the aid of using placing malicious hyperlinks into messages. To inspire customers to click on at the malicious hyperlinks, the messages are generally worded in a manner that inspires a feel of urgency or intrigue. Clicking at the hyperlink triggers the down load of ransom ware, which encrypts the gadget and holds your records for ransom. Current technique makes use of CNN to stumble on Ransom ware malware.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

A convolution neural community has a couple of hidden layers that assist in extracting information. The four important layers in CNN are:

- *Convolution Layer:* This is step one withinside the procedure of extracting treasured features. A convolution layer has numerous filters that carry out the convolution operation.
- *ReLUlayer:* It stands for the rectified linear unit. Once the characteristic maps are extracted, the subsequent step is to transport them to a ReLU layer. ReLU plays an element-sensible operation and introduces non-linearity to the community, and the generated output is a rectified characteristic map. The authentic url is scanned with a couple of convolutions and ReLUlayers for finding the features.
- *Pooling Layer:* Pooling is a down-sampling operation that reduces the dimensionality of the characteristic map. The rectified characteristic map now is going thru a pooling layer to generate a pooled characteristic map. The pooling layer makes use of diverse filters to pick out extraordinary elements of the url like length, characters area etc. The subsequent step withinside the procedure is known as flattening. Flattening is used to transform all of the resultant 2-Dimensional arrays from pooled characteristic maps right into a unmarried lengthy non-stop linear vector. Flattening is used to convert all the resultant 2-Dimensional arrays from pooled feature maps into a single long continuous linear vector. The flattened matrix is fed as input to the fully connected layer to classify the URL.

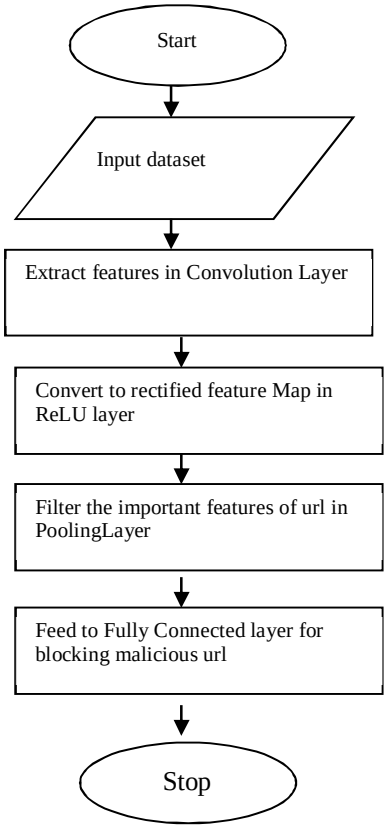


Figure 1 : Filtering malicious url using CNN

The major steps in recognizing malicious URL are:

1) The malicious url dataset are fed to the convolutional layer that plays the convolution operation
2) It effects in a convolved map .The convolved map is carried out to a ReLU characteristic to generate a rectified function map
3) The url is processed with more than one convolutions and ReLU layers for finding the capabilities
4) Different pooling layers with numerous filters are used to perceive precise elements of the url.
5) The pooled function map is flattened and fed to a completely related layer to get the very last output.

The CNN model comprises of following major modules:

- *Vector Representation Module:* This module represents the enter URL series as a appropriate vector to facilitate the following process. In the beginning, URL vector illustration is initialized randomly. The module extracts the segment statistics from the URL and the character-stage statistics from the word. The statistics extracted may be utilized in next education to attain the maximum suitable vector expression of the URL, after which the vector expression is inputted into the following convolutional layer..
- *Convolution Module:* This module extract capabilities mechanically from the enter facts. The URL is inputted into the enter layer, and it's far transformed to a appropriate vector expression withinside the embedding layer. Then, the primary convolution layer begins offevolved to extract capabilities. After the facts are outputted from the convolutional layer, the facts tensor measurement is compressed with the aid of using the folding layer after which inputted to the pooling layer for dynamic pooling. After numerous rounds of convolution-folding-pooling, the facts are sooner or later inputted into the absolutely related layer for education, and the end result is sooner or later outputted from the output layer.
- *URL Block Extraction:* After passing via the embedding layer, the second one facts department is merged with the primary facts department, and the merged end result is inputted to the absolutely related layer for education. When the block extraction module extracts extraordinary fields, it could separate the top-stage area call or countrywide area call from the URL string.
- *Detection of Malicious URL:* In the absolutely related layer, the capabilities are extracted mechanically with the aid of using the convolutional neural community and extracted artificially from the URL field. The detection version can successfully make use of vital statistics with inside the URL, which include top-stage domains and countrywide domains, to gain better accuracy and recall. Accuracy is vital, specifically for detecting models, due to the fact if the accuracy is low, regular net pages can be labelled as malicious web sites and

V. RESULTS AND DISCUSSION

The current methodology uses a third-party dataset that is a amalgamation of the following malicious URL databases.

1) Benign: Custom automated web scraping of Alexa Top 1M with recursive depth of scraping of level 1

2) Malicious: Various blacklists, openphish, phishtank, Public GitHub

There are 2 output classification labels, benign or malicious. This work contains a simple proxy server that can work with browsers to detect malicious URLs. It as a filter for malicious URLs. It works with both HTTP and HTTPs requests. In your browser settings (firefox, for example), set the proxy settings to 127.0.0.1:8080.

Figure 2:Proxy Settings Firefox

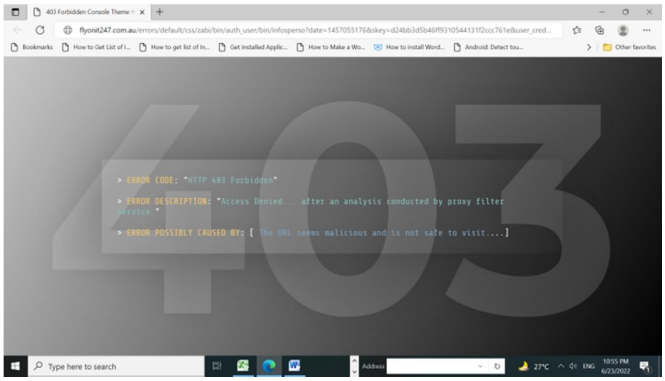


Figure 3: Malicious URL Detected

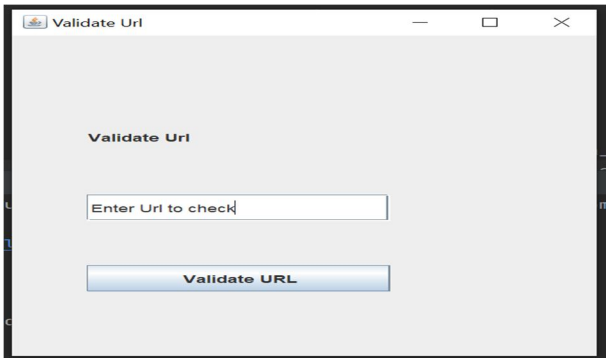


Figure 4: Feed the Url for Detection Process

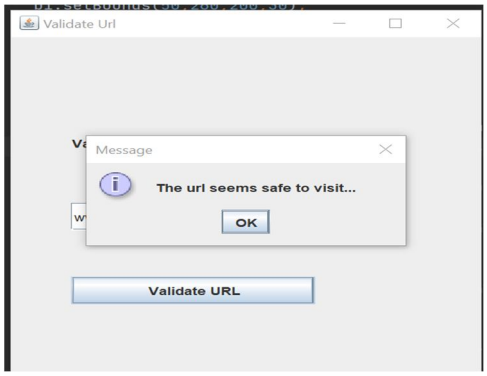


Figure 5: Validate if the url is safe to visit

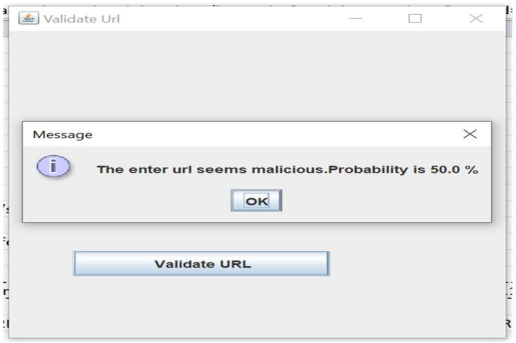


Figure 6: Probability score shown if url is malicious

Now you can try to navigate to a misspelled URL in the browser navigator and see the proxy at work. The performance of the neural network model is calculated as per mentioned parameters:
Confusion Matrix: A tabular summary of True/False Positive/Negative prediction rates.

Prediction Classes	True Values for Classes Malicious	True Values for Classes Non Malicious
Malicious	43	0
Non Malicious	18	42
Total	71	42

Table 1 : Confusion Matrix for URL Prediction

We calculate accuracy by dividing the number of correct predictions (the corresponding diagonal in the matrix) by the total number of samples.

Accuracy = Sum of diagonal values / Total Sample ----(1)

Accuracy = 85 /113= 75%

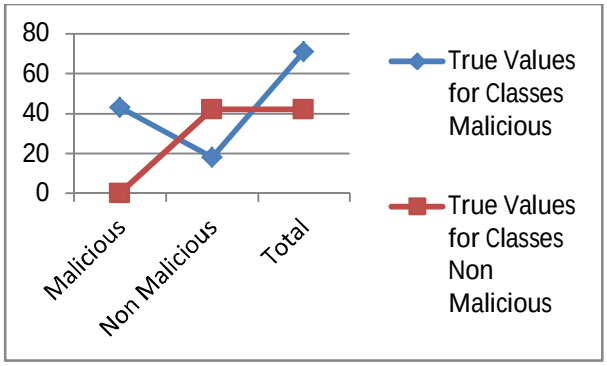


Figure 5: Classification Accuracy

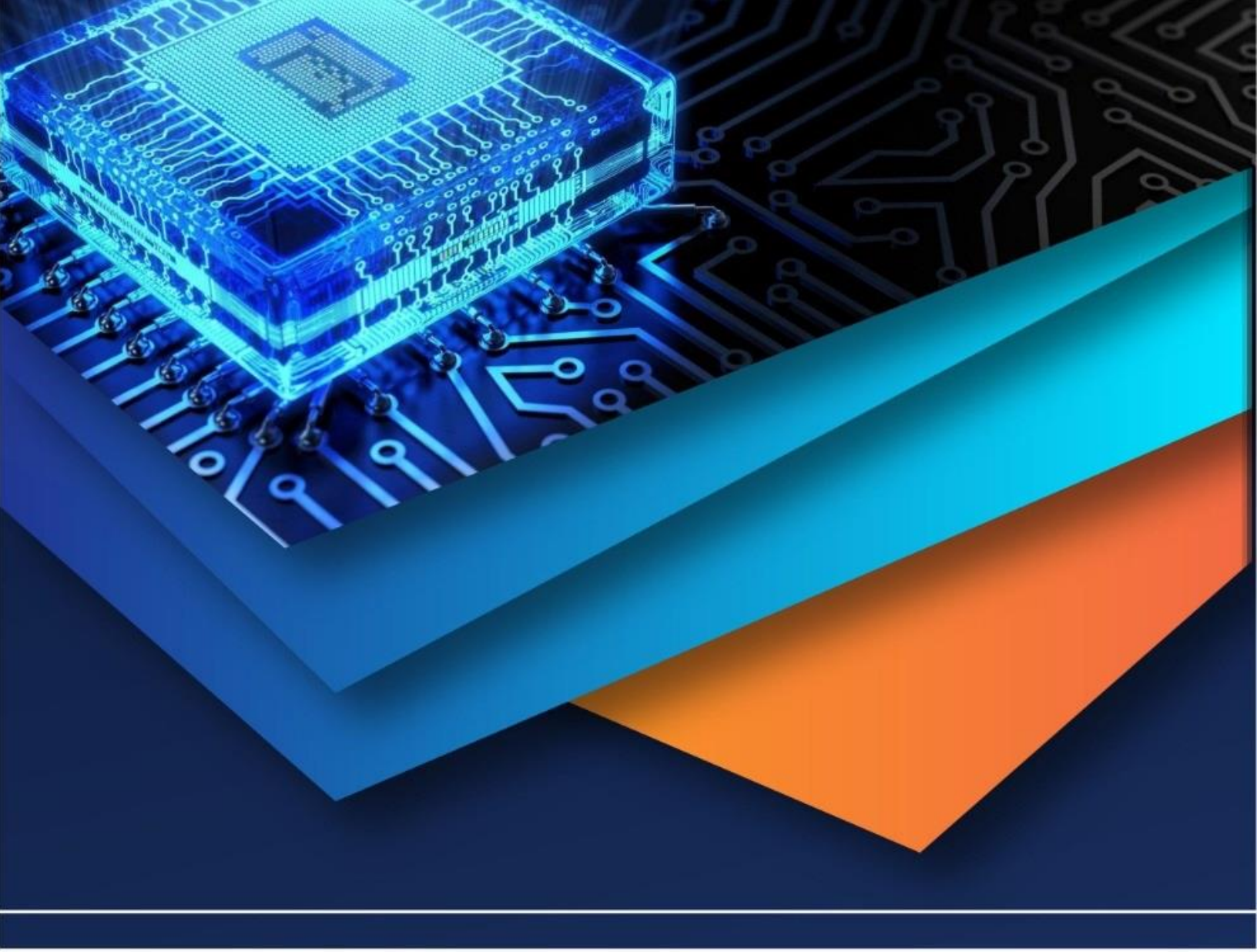
VI. CONCLUSION

Ransom ware assaults are exceptionally well known to the assailants as they are made or delivered income for aggressors. Additionally Ransom ware assault becomes most impressive danger to individual and associations as they stop the working of frameworks by assaulting and encoding records or frameworks. The current work attempts to detect the malicious url which are major medium through which a ransom ware spreads using CNN. The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any human supervision. The future work aims to explore the uses of CNN in detection of ransom ware plugins.

REFERENCES

[1] Anti-Phishing Working Group (APWG) Phishing Activity Trends Report, 1st Quarter 2021, Anti-Phishing Working Group, Inc. (2021), https://docs.apwg.org/reports/ apwg_trends_report_q1_2021.pdf.
[2] Cooper, M., Levy, Y., Wang, L. and Dringus, L. Heads-up! An alert and warning system for phishing emails, Organizational Cybersecurity Journal: Practice, Process and People, 2021, 1–22.
[3] Akdemir, N. and Yenal, S. How Phishers Exploit the Coronavirus Pandemic: A Content Analysis of COVID-19 Themed Phishing Emails, SAGE Open, 11(3), 2021, 1– 14,doi: 21582440211031879.
[4] Mourtaji, Y., Bouhorma, M., Alghazzawi, D., Aldabbagh, G.andAlghamdi, A. Hybrid Rule-Based Solution for Phishing URL Detection Using Convolutional Neural Network, Wireless Communications and Mobile Computing, 2021, 1–24.
[5]] Mourtaji, Y., Bouhorma, M., Alghazzawi, D., Aldabbagh, G.andAlghamdi, A. Hybrid Rule-Based Solution for Phishing URL Detection Using Convolutional Neural Network, Wireless Communications and Mobile Computing, 2021, 1–24
[6] Evans, K., Abuadbba, A., Ahmed, M., Wu, T., Johnstone, M., and Nepal, S. RAIDER: Reinforcementaided Spear Phishing Detector, arXiv preprint arXiv:2105.07582, 2021, 1–19.
[7] D.Warren Fernando, N.Komninos, T.Chen,A Study on the Evolution of Ransomware Detection Using Machine Learning and DeepLearning Techniques, www.mdpi.com/journal/iot,Dec .2020
[8] HR, M. G., Adithya, M. V. and Vinay, S. Development of anti-phishing browser based on random forest and rule of extraction framework, Cybersecurity. 3(1), 2020, 1–14.
[9] El Aassal, A., Baki, S., Das, A. and Verma, R. M. An indepth benchmarking and evaluation of phishing detection research for security needs, IEEE Access, 8, 2020, 22170– 22192, doi: 10.1109/ACCESS.2020.2969780.
[10] Mohammada, G. B., Shitharthb, S. and Kumarc, P. R. Integrated Machine Learning Model for an URL Phishing Detection, International Journal of Grid and Distributed Computing, 14(1) , 2020, 513–529.
[11] DanielGibert,CarlesMateu,JordiPlanes," The rise of machine learning for detection and classification of malware: Researchdevelopments, trends and challenges",2020, Journal of Network and Computer ApplicationsVolume 153, 1 March 2020, 102526.
[12] Bander Ali Shah Al-rimy , Crypto-ransomware early detection model using novel incremental bagging with enhanced semi-random subspace selection,Future Generation Computer SystemsVolume 101Issue CDec 2019 pp 476–491https://doi.org/10.1016/j.future.2019.06.005.
[13] SajadHomayoun, AliDehghantanha, DRTHIS: Deep ransomware threat hunting and intelligence system at the fog layer,Future Generation Computer Systems Volume 90, January 2019, Pages 94-104
[14] M.Ijaz, M.Hanif Durad, M.Ismail , Static and Dynamic Malware Analysis Using Machine Learning, 2019 IEEE.
[15] Sahingoz, O. K., Buber, E., Demir, O. and Diri, B. Machine learning based phishing detection from URLs, Expert Systems with Applications, 117, 2019, 345–357, doi: https://doi.org/10.1016/j.eswa.2018.09.029.

[16] Zhu, E., Chen, Y., Ye, C., Li, X. and Liu, F. OFS-NN: an effective phishing websites detection model based on optimal feature selection and neural network, IEEE Access. 7, 2019, 73271–73284, doi: 10.1109/ACCESS.2019.2920655.
[17] Mao, J., Bian, J., Tian, W., Zhu, S., Wei, T., Li, A., and Liang, Z. Phishing page detection via learning classifiers from page layout feature, EURASIP Journal on Wireless Communications and Networking, 1, 2019, 1–14, doi: https://doi.org/10.1186/s13638- 019-1361-0.
[18] Adebowale, M. A., Lwin, K. T., Sanchez, E. and Hossain, M. A. Intelligent web- phishing detection and protection scheme using integrated features of Images, frames and text, Expert Systems with Applications. 115, 2019, 300– 313.
[19] Patil, D. R., Patil, J. B. Malicious web pages detection using feature selection techniques and machine learning, International Journal of High Performance Computing and Networking., 14(4), 2019, 473–488., doi: 10.1504/IJHPCN.2019.102355.
[20] T. Boczan. (Jul. 2019). The Evolution of GandCrab Ransomware.Accessed:.[Online]. Available: https://www.vmray.com/cyber-security-blog/gandcrab-ransomware-evolution-analysis/
[21] Zabihimayvan, M., Doran, D. and Solouk, V. Fuzzy Rough Set Feature Selection to Enhance Phishing Attack Detection, arXiv preprint: 1903.05675. (2019) 1–6.
[22] Babagoli, M., Aghababa, M.P, M.P. and Solouk, V. Heuristic nonlinear regression strategy for detecting phishing websites, Soft Computing. 23(12), 2019, 4315– 4327.
[23] Sahoo, D., Liu, C., Hoi, S. C., and Solouk, V. Malicious URL Detection using Machine Learning: A Survey, arXiv preprint arXiv:1701.07179. 2019, 1–37.
[24] Ding, Y., Luktarhan, N., Li, K. and Slamu, W. A keywordbased combination approach for detecting phishing web pages, Computers & Security. 84, 2019, 1–6, doi:10.https://doi.org/10.1016/j.cose.2019.03.018.
[25] Aurélien Palisse, Data Aware Defense (DaD): Towards a Generic and Practical Ransomware Countermeasure, TAMIS - Threat Analysis and Mitigation for Information Security Inria Rennes – Bretagne Atlantique , IRISA-D4 - LANGAGE ET GÉNIE LOGICIEL,2018
[26] S. K. Shaukat and V. J. Ribeiro, ‘‘RansomWall: A layered defense system against cryptographic ransomware attacks using machine learning,’’ in Proc. 10th Int. Conf. Commun. Syst. Netw. (COMSNETS), Jan. 2018,pp. 356–363.
[27] Omar M. K. Alhawi, James Baldwin, and Ali Dehghantanha, “Leveraging Machine Learning Techniques for Windows Ransomware Network Traffic Detection”, Springer International Publishing AG, part of Springer Nature 2018, Cyber Threat Intelligence,Advances in Information.Security 70, https://doi.org/10.1007/978-3-319- 73951-9_5.
[28] Patil, D. R., Patil and J. B. Malicious URLs detection using decision tree classifiers and majority voting technique, Cybernetics and Information Technologies, 18(1), 2018, 11–29, doi: https://doi.org/10.2478/cait-2018-0002.
[29] Dae-Youb Kim, Geun-Yeong Choi, and Ji-Hoon Lee, “White List-based Ransomware Real-time Detection and Prevention for User Device Protection”, 2018 IEEE International Conference on Consumer Electronics (ICCE), 978-1-5386-3025-/18/$31.00 ©2018 IEEE
[30] Bander Ali Saleh Al-rimy(&), Mohd Aizaini Maarof, and Syed Zainuddin Mohd Shaid, “A 0-Day Aware Crypto-RansomwareEarly Behavioral Detection Framework”, Springer International Publishing AG 2018, Recent Trends in Information and Communication Technology, Lecture Notes on Data Engineering and Communications Technologies 5, DOI 10.1007/978-3- 319-59427-9_78.
[31] M. AL-Hawawreh, N. Moustafa, E. SitnikovaIdentification of malicious activities in industrial internet of things based on deep learning modelsJournal of Information Security and Applications, Journal of Information Security and Applications, 41 (2018).
[32] Niakanlahiji, A., Chu, B. T. and Al-Shaer, E. PhishMon: A Machine Learning Framework for Detecting Phishing Web pages, In : IEEE Int. Conf. Intelligence and Security Informatics (ISI), (Miami, FL, USA, 2018), pp. 220–225.
[33] Yuan, H., Chen, X., Li, Y., Yang, Z. and Liu, W. Detecting Phishing Websites and Targets Based on URLs and Webpage Links, In: Int. Conf. Pattern Recognition (ICPR), (Beijing, China, 2018), pp. 3669–3674.
[34] Patil, D. R.Patil and J. B. Feature-based Malicious URL and Attack Type Detection Using Multi-class Classification, The ISC International Journal of Information Security (ISeCure). 10(2), 2018, 141–162, doi:10.22042/ISECURE.2018.0.0.1.
[35] Mishra, A. and Gupta, B. B. Intelligent phishing detection system using similarity matching algorithms, International Journal of Information and Communication Technology. 12(1-2), 2018, 51–73.
[36] Md Mahbub Hasan, RansHunt: A support vector machines based ransomware analysis framework with integrated feature set,December 2017 DOI:10.1109/ICCITECHN.2017.8281835 Conference: 2017 20th International Conference of Computer and Information Technology (ICCIT).
[37] Cutting the Gordian Knot: A Look Under the Hood of Ransomware Attacks - Amin Kharraz , William Robertson , Davide Balzarotti , Leyla Bilge , EnginKirda,2017
[38] Arab, M., and Sohrabi, M. K. Proposing a new clustering method to detect phishing web- sites, Turkish Journal of Electrical Engineering and Computer Sciences. 25(6), 2017, 4757–4767
[39] B Jyothi Kumar, Naveen H, B Praveen Kumar, Sai Shyam Sharma,Jaime Villegas,”Logistic Regression For Polymorphic Malware Detection Using ANOVA F-Test,2017 International Conference on Innovations in information Embedded and Communication Systems (ICIIECS).
[40] B. Athiwaratkun, J.W. StokesMalware classification with lstm and gru language models and a character-level cnn2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (March 2017).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY