Journal of Computing, Science &Technology
https://focjournal.unidel.edu.ng/
editorjcst@unidel.edu.ng
ISSN:......Print ISSN: .... Online

Malicious Domain Names Detection using Deep-Learning Classifiers

Egwali Annie .O[1], and Ekhator Roland O[2].

[1,2]Department of Computer Science, Faculty of Physical Sciences. University of Benin, PMB 1154, Benin City, Nigeria.

annie.egwali@uniben.edu[1], osagie.ekhator@physci.uniben.edu [2]

Corresponding Author's Email: *annie.egwali@uniben.edu*

ABSTRACT

Article Info

Date Received: 14-01-2024
Date Accepted: 11-04-2024

The strategies growth of innovative technologies used for online services in the global economic space brings vulnerabilities to security breaches. The implication of these vulnerabilities created a level playing field for cyber-attacks to flourish, with assailants constantly adapting new nefarious methods to compromise information and deceive naïve users of the cyberspace. Despite the amazing and numerous anti-phishing approaches and solutions, the increasing rate caused by malicious domain name system attacks such as spam, phishing and malware could be attributed to the dynamism in the approaches used by cybercriminals to counterfeit the techniques. To address these issues, many cyber security researchers have switched their focus to machine learning-based methodologies for malicious DNS detection. In this paper, we incorporate the usage of machine-based model to detect the dynamism of malicious DNS by exploring the intricate feature extraction capabilities inherent benefits offered by the models. A customized web Crawler was implemented to extract URL attributes for model extraction on datasets packets compromising of Malware (182,266) that are designed to disrupt, damage, steal or gain unauthorized access to the system, spam (61,046) unsolicited messages, sent in bulk to *a large list of recipients, phishing (95,492) to deceive users and extract vital information like passwords or financial information and benign (6,907,719) packets determined by the system to be harmless or not a threat. Furthermore, we also leveraging the capabilities of the various models to demonstrate the feasibility of identifying malicious DNS through k-holdout approach in order to assess their performance of the integrated model with an accuracy of 89.9%. Our experiment is based on both active and passive DNS analysis.*

Keywords:
Malicious Domain Detection, DNS, Cyber Security, Deep Learning, Domain Generation.

1.0 INTRODUCTION

Recently, widely reported cybercrimes perpetrated through malicious domain used for online services and businesses in the global digital space brings along vulnerabilities to network security. Lots of diverse domain generation algorithms are used by attackers comprehensively to generate a massive amount of random domain names to evade detection [1]. A vulnerable network paves the way for hackers to manipulate and violate system infrastructure [2]. The rise of these vulnerabilities has created a level ground for cyber-attacks to bloom. Often time domains that have recently engaged in suspicious acts are Black listed while the white lists include well-known and trustworthy domain names ([3], [4], [5]). This scenario is only used for broadly distributed infections rather than targeted ones. Attackers implant malicious programs (code) through the network vulnerabilities into host which grants them access to remotely control the host ([6], [7], [9]). The affected host then issue resolution requests, using a large number of nonexistent domain names randomly generated by the domain generation algorithm (DGA), a program that generates a large list of domain names and provide malware with new domains to evade security countermeasures in a short time ([10], [11]). However, malicious domain refers to the domain registered by attackers which not well and remain active for a short period to avoid detection [12]. Domain names are listed in public lists based on their behavior and they objectively evaluate the reputation of a domain name. Malicious domain name (MDN) attack is a persisting problem in domain name system (DNS). Several researchers assert the fact that more work should be done and direction should be focused on the attacks especially botnets [13]. Some selected problems are limitations of conventional approaches used by many for malicious domain detection. According to [14], these approaches are blacklisting of domain names, analysis of the network traffic, dissection of the webpage content, DNS traffic analysis, and analysis of salient lexical features. The non-consideration of the domain name and the DGA data for computing the maliciousness of the URL results in a lack of precision. Hence, they advised effective mechanisms for malicious domain detection to help improve the precision of malicious URL detection using algorithm features selection and dynamic machine learning models that can act passively and actively in detecting malicious

DNS; making use of Deep Learning algorithms to detect malicious accounts based on domain names to blacklisting associated Ips.

This paper aim therefore to designed a system of detecting DNS using industry proven and process model methodology called Cross Industry Standard for Data Mining (CRISP-DM), and trained the proposed model using DGA dataset and to compare and validate the model performance (Accuracy, Precision, F1-score and Recall or regression metrics), using K-holdout approach.

1.1. Related Works

Various computational models have been investigated for detecting the malicious domain names in establishing protected DNS territory. Some of the recent studies were discussed. The verification of malicious DNS using multiple features can also be described in different dimension, in order to distinguish legitimate and malicious domains. A MDN detection technology was proposed using a passive domain name analysis method, and a technique called EXPOSURE to train and monitor the domain name traffic of a commercial ISP [11]. Several listed features were extracted by the researchers including the domain name lifetime, period similarity number of accesses; number of IPs parsed, whether IP is shared by other domain names, digital symbol ratio and length of longest meaningful substring. However, a classifier was constructed using decision tree algorithm.

A botnet detection algorithm was identified in [15] based on DNS traffic features using Power Spectral Density (PSD) testing technology which detects MDNs by analyzing malicious behavior within large volumes of DNS traffic. In 2016, Woodbridge et. al. [16] presented a DGA classifier that influence long short-term memory (LSTM) networks for real-time production of DGA’s without the need for contextual information or manually created features. The experimental was significantly performed best than some state-of-the-art techniques. **Vinayakumar** et. al. [17] obtained the data for the study of Local DNS records and investigated the performance of various DL algorithms such as RRN, LSTM, and other approaches in which LSTM showed better in identifying malicious DNS requests. Zang et. al. [18] proposed a MDNs detection algorithm based on AGD (Algorithmically generated domain) by using cluster correlation to identify the names generated by a domain generation algorithm or its variants. Various features such as TTL, the distribution of IP Addresses; WHOIS features, and historical information from the domain names in each cluster were extracted and the Support Vector Machine (SVM) algorithm was used to identify the MDNs.

Deep Learning (DL) approaches for the recognition of fraudulent domain names, [19] extracted textual characteristics from domain names by passing them to LSTM and bidirectional LSTM. Palaniappan et. al. [20] introduced features selection are; blacklist domain names features, DNS based features, web-based features and lexical features to identify malicious domain through features extracted from domain names. Mvula et al. [21] proposed a machine model that uses few features to classify COVID-19 related DNs as legitimate or malicious. Their result confirmed that a smallest set of features (lexical features) extracted from domain names had made the model to achieved a high scores. Suliang et. al. [22] aim to use a new metric to evaluate real unbalanced traffic data. Their experimental result shows that the level of the precision model and the value of the area under the curve (AUC) reach a certain maximum height. Sachan et. al. [23] develop a system that detects the feasibility of MDN account in relation to block chain so as to know whether it is malicious or not. They also use numerous features such as DN string based, DNS query based, DNS graph based and temporal aspect based extracted from domain names.

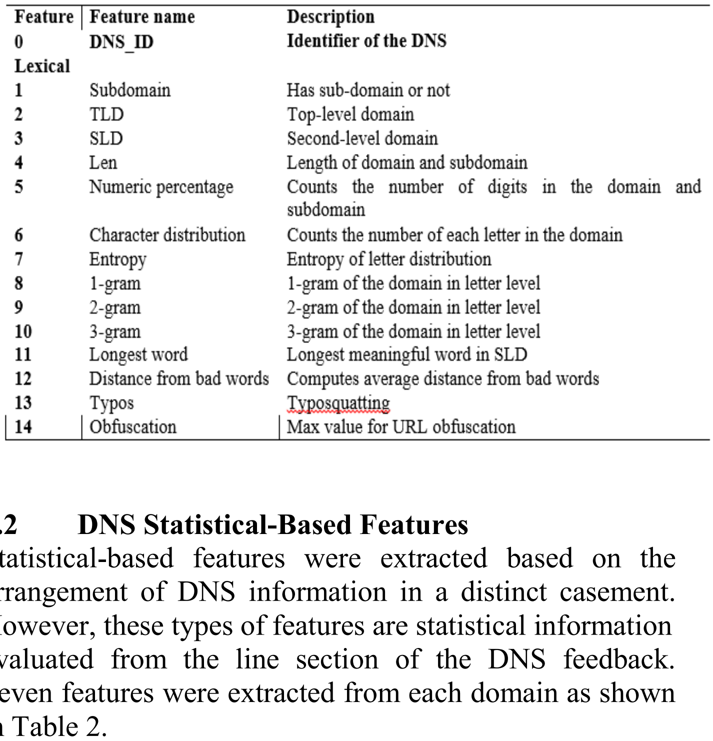
2.0 METHODOLOGY

Our proposed approach is based on three features which we extract from domain name online repository categorizing them into three groups; lexical-based, DNS statistical-based and third party-based features.

2.1 Lexical-based features

The lexical features ensure that MDNs can be detected using various features. In this work, we have extracted fourteen features from each domain details as represented in table 1:

Table 1: List of DNS-based features



Feature	Feature name	Description
0	**DNS_ID**	**Identifier of the DNS**
Lexical		
1	Subdomain	Has sub-domain or not
2	TLD	Top-level domain
3	SLD	Second-level domain
4	Len	Length of domain and subdomain
5	Numeric percentage	Counts the number of digits in the domain and subdomain
6	Character distribution	Counts the number of each letter in the domain
7	Entropy	Entropy of letter distribution
8	1-gram	1-gram of the domain in letter level
9	2-gram	2-gram of the domain in letter level
10	3-gram	3-gram of the domain in letter level
11	Longest word	Longest meaningful word in SLD
12	Distance from bad words	Computes average distance from bad words
13	Typos	Typosquatting
14	Obfuscation	Max value for URL obfuscation

2.2 DNS Statistical-Based Features

Statistical-based features were extracted based on the arrangement of DNS information in a distinct casement. However, these types of features are statistical information evaluated from the line section of the DNS feedback. Seven features were extracted from each domain as shown in Table 2.

Table 2: List of DNS statistical-based features

DNS statistical		
1	Unique country	The number of distinct country names in the window
2	Unique ASN	The number of distinct ASN values in the window
3	Unique TTL	The number of distinct TTL values in the window
4	Unique IP	The number of distinct IP values in the window
5	Unique domain	The number of distinct domain values in the window
6	TTL means	The average TTL in the window
7	TTL variance	The variance of TTL in the window

2.3 Third-party Features

Twelve features of the third party were extracted from two sources; WHOIS and Alexa rank. They contain the biographical properties of a domain as shown the table 3.

Table 3: Third-party Features Extracted

1	Domain name	Name of the domain
2	Registrar	Registrar of the domain
3	Registrant name	The name of the domain has been registered
4	Creation date time	The date and time the domain created
5	Emails	The emails associated with a domain
6	Domain age	The age of a domain
7	Organization	What organization it is linked to
8	State	The state the main branch is
9	Country	The country where the main branch is
10	Name server count	The total number of name servers linked to the domain
11	Alexa rank	The rank of the domain by Alexa
12	Status	The class of the DNS-Benign or Malicious

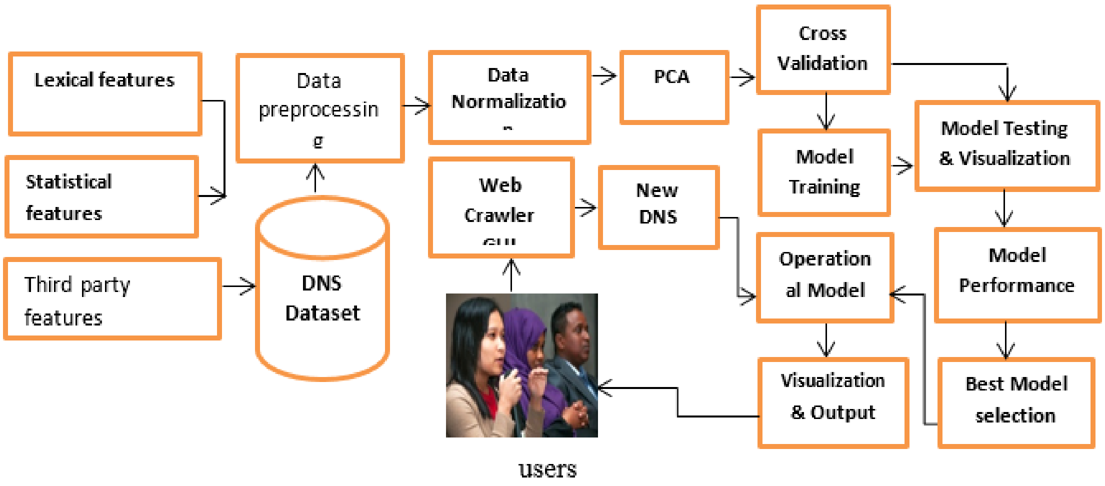


Figure 1: Proposed Malicious DNS prediction system

The software methodology adopted in this study is the Cross Industry Standard Process for Data Mining (CRISP-DM). The CRISP-DM is both an industry-proven methodology and a process model. As an industry methodology, it provides a concrete description of typical project stages tasks associated with each stage as well as the details of the interrelationship between the various tasks. As a process model, it provides a sketch that shows the data mining. In this study, CRISP-DM undergoes five stages include; (i) Business and data understanding, (ii) Data preparation, (iii) Modeling, (iv) Evaluation, and (v) Deployment. Figure 1 shows the proposed architecture, adopting the CRISP-DM methodology.

Table 4: Statistics of the domains dataset

Category	Original domains	Domains processed	Packets processed	Size in megabytes
Malware	26,895	9,432	182,266	2.7
Spam	8,254	1,976	61,046	2.4
Phishing	16,307	12,586	95,492	2.8
Benign	988,667	500,000	6,907,719	266

2.4 Data Preparation (DNS Dataset Description)

The CICBellDNS2021 was extracted at the Canadian Institute for Cyber security. The dataset compromises malware, spam, phishing, and benign URLs stored in comma-separated value (CSV) files (See table 4). The Correlation Matrix Analysis and Principal Component Analysis were then introduced to determine the relationship that exist among features in the dataset so that the most highly correlated features with predictor will be considered relevant for model building. The study was implemented using an Intel Core i7 processor with 8GB of RAM and a 300GB hardisk. Python was the chosen language and Keras and Sci-Learn were the standard libraries for building models. The implementation was conducted in two phases; exploratory data analysis (EDA) for future engineering that consists of descriptive analysis with data normalization and model development that consists of training, validation and fine-turning of the various models used. The study also applied data modeling (design) such as data flow data (DFD), entity relationship diagram (ERD) and class diagram (CD) to show the links between data points and structure within the models.

3.0 RESULTS AND DISCUSSION

Figure 2 shows the datasets consisting of 16 attributes. Some features were not correlated due to inter-feature correlation that was performed very highly in correlation to one another, were identified and necessary to remove to avoid multicollinearity leading to model imprecision. These features were dropped to improve model performance. Hence, a total of 16 predictor features and one target feature (status) were used in the dataset for the model-building phase. For the model to perform better, the epoc value was increased from 5 to 10 as shown in figure 2. Figure 3 shows that all 16 features extracted from the 16 features of the data preprocessing significantly contribute to the variance in the dataset, hence they were used for model training with the help of Epoc values to determine the number of iterative or training that the model will perform. Figure2: Correlation Matrix for dataset

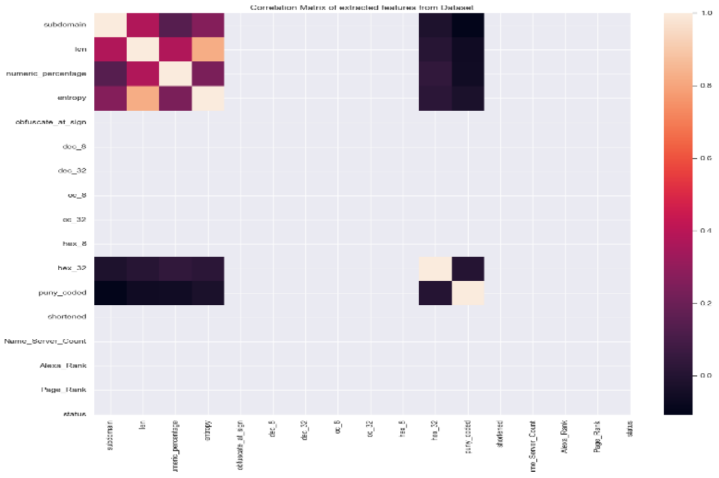


Figure 2: Correlation Matrix for dataset

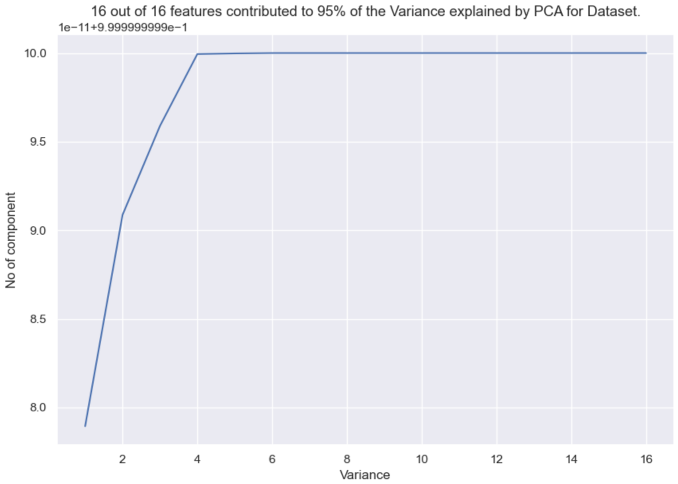


Figure 3: PCA diagram at 0.95% for dataset

3.1 Model Evaluation and Performance

Using the testing dataset, the trained models were evaluated. The models predict the labels of the test samples, which are then compared to the true labels to determine their performance. The evaluation of the various models were conducted using several performance and validity metrics. Some existing system used features like DNS-base, web-based, blacklisting and lexical-based features to classify domain name ([20, [4], 16]) whereas our proposed approach is based on three features, the lexical-based, DNS-statistical and third party features, extracted from a domain repository online.

Furthermore some existing system used logistic regression classifier model to classify unlabeled datasets of domain names and got an average accuracy rate of 60%, but our proposed system used classification and regression metrics using 5-fold cross validation classification and regression metrics. Table 5 shows the model's 5-fold cross validation classification metrics. In the EL models categories, RF had the highest accuracy of 89.9%. In ML models, DT had the highest accuracy of 86.9%. In the DL models, GRU had the highest accuracy score of 77.2%. Hence, RF (Ensemble Learning) had the highest accuracy score of 89.9% at 5-fold cross validation approach and was considered appropriate for the best detection model for predicting malicious DNS.

Table 2 Model Classification Metric using 5-fold cross validation approach

Model	Accuracy	Recall	Precision	F-Score
DT	86.9	86.0	89.7	87.8
LogR	65.5	67.6	66.0	66.8
SVM	65.5	67.7	65.7	66.7
ANN	66.0	71.5	58.5	64.3
KNN	85.2	91.1	79.6	85.0
RF	89.9	89.2	91.9	90.5
Xgboost	83.5	86.7	81.1	83.8
MLP	73.1	72.3	80.0	75.6
DNN	72.6	74.4	72.9	73.6
GRU	77.2	77.3	80.9	78.6
LSTM	73.8	75.8	75.8	75.1

Table 5: Model Regression Metric using 5-fold cross validation approach

Algorithm	R^2	MSE	RMSE
DT	0.48	0.13	0.36
LogR	-0.38	0.34	0.59
SVM	-0.38	0.34	0.59
ANN	-0.36	0.34	0.58
KNN	0.41	0.15	0.38
RF	0.60	0.10	0.32
Xgboost	0.34	0.16	0.41
MLP	-0.08	0.27	0.52
DNN	-0.10	0.27	0.52
GRU	0.08	0.23	0.48
LSTM	-0.05	0.26	0.51

For the root mean square error, we choose the best model by identifying the least value in the evaluation metrics. Hence, in the EL models categories; RF had the lowest

RMSE of 0.60. In ML models, DT had the lowest RMSE of 0.48. In the DL models, GRU had the lowest RMSE of 0.08. Hence, GRU has the least error rate and it is considered best for the detection of malicious DNS in terms of reduce error rate of detecting malicious DNS. The implementation of RF allows for relatively quick and efficient malicious detection, making it suitable for real-time applications such as email security systems or web browsers. This is also applicable on mobile responsive website that is capable of adapting its content based on the device it is being observed on [24].

4.0 CONCLUSION AND FUTURE WORK

The study's benefits lie in its high accuracy, improved DNS detection, adaptability, practical implementation, and potential for integration. These implications can significantly enhance users' protection against malicious DNS attacks and contribute to the advancement of cyber security defenses. This study demonstrates the development of three modeling approaches for detecting malicious DNS with the ability to quantify the uncertainty in the prediction or detection. The modeling approach entails various classifiers which were evaluated to the best performance. Also, the models were evaluated using 5-fold cross-validation to ensure that they were exposed to all data and to detect potential over-fitting a procedure frequently used within. The models obtained outperformed a better fit of the experimental data with an accuracy of 89.9%.

In our future work, we look forward to extending its benefits to a broader user base (web browser plugin) and strengthening overall cyber security measures as well as to use another combine models to optimize Ensemble Learning classifier (Random Forest) to a real-time malicious DNS detection by implementing it as domain specific application for a better accuracy performance.

REFERENCES

[1] Liu W., Ma X., Wang H. and Wu Z. (2023). A generation method of malicious domain name training data based on generating adversarial network. Journal of Lanzhou University of Technology, 2023, 49(6): 100-106.

[2] Li, K., Yu, X., and Wang, J. (2021) A Review: How to Detect Malicious Domains. In *Advances in Artificial Intelligence and Security: 7th International Conference, ICAIS Dublin, Ireland, July 19-23, 2021, Proceedings, Part III 7* (pp. 152-162). (2021) Springer International Publishing.

[3] Hamroun, C., Amamou, A., Haddadou, K., Haroun, H., and Pujolle, G.(2022.) A review on lexical based MDN detection methods. *6th Cyber Security in Networking Conference (CSNet)* (pp. 1-7). IEEE.

[4] Egwali A. O. and Alile S. O. (2020). A Casual Network Based System For Predicting Multi-Stage Attack with Malicious IP. International Journal of Academic Multidisciplinary Research 4 (5), 1-8.

[5] Alile S. O. and Egwali A. O. (2020). A Bayesian Belief Network Model For Detecting Multi-stage Attacks With Malicious IP Addresses. I.J. Wireless and Microwave Technologies, 2, 30-41

[6] Bilge L., Kirda, Kruegel E. and Balduzzi C. (2011). EXPOSURE. Finding Malicious Domains Using Passive DNS Analysis. In Proceedings of the 18th Network and Distributed System Security Symposium, San Diego, CA, USA, 6 February 2011; Internet Society: Reston, VA, USA, PP. 1-17.

[7] Hong Z., Zhaobin C., Guangbin B. and Xiangyan Z. (2019). MDNs detection algorithm based on N-Gram. Journal of Computer Networks and Communications, Volume 2019, Article ID 4612474, page 9 Publish 3 February 2019. Guest Editor: Saman S. Chaeikar.

[9] Halgamuge M. N. (2022). Estimation of the success probability of a malicious attacker on blockchain-based edge network. *Computer Networks*, *219*, 109402.

[10] Mitsuhashi R., Jin Y., Iida K., Shinagawa T. and Takai Y. (2023). Detection of dga-based malware communications from doh traffic using machine learning analysis. In 2023 IEEE 20th Consumer Communications & Networking Conference (CCNC), pages 224–229. IEEE.

[11] Ayub, M. A., Smith, S., Siraj, A., & Tinker, P. (2021, June). Domain Generating Algorithm based Malicious Domains Detection. In *2021 8th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2021 7th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)* (pp. 77-82). IEEE.

[12] Maroofi, S., Korczyński, M., Hesselman, C., Ampeau, B., & Duda, A. (2020, September). COMAR: classification of compromised versus maliciously registered domains. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)* (pp. 607-623). IEEE.

[13] Ma D. J., Zhang S., Kong F. and Fu Z. (2021). MDN Detection Based on Doc2vec and Hybrid network. In IOP conference series: and Environmental Science. IOP publishing: Bristo, UK, Volume 693 [Cross Ref]

[14] Palaniappan, G., Sangeetha, S., Rajendran, B., Goyal, S., & Bindhumadhava, B. S. (2020). Malicious domain detection using machine learning on domain name features, host-based features and web-based features. *Procedia Computer Science*, *171*, 654-661.

[15] Kwon J., Lee J., Lee H., and perring A. (2016). "PsyBog:Psy Bog: a scalable botnet detection method for large scale DNS Traffic", Computer Networks vol.97, pp.48-73.

[16] Woodbridge, J.; Anderson, H.S.; Ahuja, A.; Grant, D. Predicting Domain Generation Algorithms with Long Short-Term Memory Networks. arXiv 2016, arXiv:1611.00791. [Google Scholar]

[17] Vinayakumar R. (2023). Detecting malicious domain names using deep learning approaches at scale 1367. ResearchGate. Available from: https://www.researchgate.net/figure/Summary-of-test-results-of-Data-set-1-for-classifying-domain-name-as-either-benign-or_tbl2_323982138

[18] Zang X., Gong J. and Hu X. (2018). "Detecting malicious domain name based on AGD," *Journal on Communications*, vol. 39, no. 7, pp. 15–25, 2018.

[19] BharathiB.,andBhauvana J. (2019). Domain name detection and classification using deep neural networks. In international symposium on Security in Computing and Communication; Springer: Singapore.

[20] Palaniappan, G., Sangeetha, S., Rajendran B., Goyal, S. and Bindhumadhava B. S. (2020). Malicious domain detection using machine learning on domain name features, host-based features and web-based features. *Procedia Computer Science*, *171*, 654-661.

[21] Mvula, P. K., Branco, P., Jourdan, G. V., & Viktor, H. L. (2022). COVID-19 malicious domain names classification. *Expert Systems with Applications*, *204*, 117553.

[22] Suliang, Luo., Gang, Han., An, Li., Jialiang, Peng. (2022). Detecting malicious domain names from domain generation algorithms using bi-directional LSTM network. 12455:124550R-124550R. doi: 10.1117/12.2655178.

[23] Sachan R. K., Agarwal R. and Shukla S. K. (2023). Identifying malicious accounts in blockchains using domain names and associated temporal properties, Blockchain: Research and Applications,Volume 4, Issue 3,100136,ISSN 2096-7209,https://doi.org /10.1016/j.bcra.2023. 100136.

[24] Ward C. (2017). Jump Start Responsive Web Design: Modern Responsive Solutions. SitePoint; 2nd edition. Pages 11 - 14.