Detecting Malicious Domains: A Review

Samar Abbas Mangi*[1], Samina Rajper[2], Noor Ahmed Shaikh[3], Nizamuddin Maitlo[4], Asad Hameed Soomro,[5] Waseem Ahmed Buriro[6]

Abstract: **Malicious domain names are an important and worrying indicator of cyberattacks and can pose serious risks to your privacy and property. Unwary Internet users can obtain malicious services from these domains, including spam servers, phishing sites, and command and control (C&C) servers. Therefore, developing efficient algorithms to identify tumor regions has attracted much attention and interest. Data sources and implementation strategies used by current detection technologies vary widely. In this study, we performed a comprehensive retrospective analysis of these methods and divided the data into Domain Name System (DNS) data and DGA data. Researchers must use appropriate detection techniques that match the unique characteristics of the data, because different data sources provide different data models and carry different information. Therefore, the detection method is divided into four types. For each method, we describe a general detection framework that defines the main steps and processes involved. In addition, we provide insight into the future potential of research on malicious domain detection. By examining existing methods for detecting and identifying vulnerabilities, this document contributes to the fight against the ever-changing threat of malicious domains, ultimately improving the security of Internet users worldwide.**

***Keywords*: Malicious domain, cyberattacks, phishing sites, efficient algorithms, detection technologies, Domain Name System.**

INTRODUCTION

With the continuous and rapid advancements in the Internet and information technology, network security concerns have become increasingly significant. Consequently, cyberattacks remain a persistent threat. Attackers frequently exploit the Domain Name System (DNS) due to its critical role in the Internet's infrastructure. To execute malicious activities, such as managing spam servers, hosting phishing sites, and operating command-and-control (C&C) servers, attackers leverage DNS to generate domain names. These malicious domains act as key enablers for many cyberattacks prevalent in today’s digital landscape.

The DNS forms the backbone of the modern Internet. Its primary function is to translate complex, hard-to-memorize IP addresses into simpler domain names. The domain namespace operates as a hierarchical tree structure that organizes domain names. At the apex of this structure is the root domain, symbolized by an empty label. This structure is supported by a hierarchical database containing resource records.

The domain namespace is hierarchically organized into zones that are managed by different authorities. The creation of top-level domains (TLDs) and its management are overseen by a non-profit organization called Internet Corporation for Assigned Names and Numbers (ICANN), which delegates the control of these TLDs to various registries. These TLDs are managed by different registrars per domain. Nevertheless, the DNS infrastructure is not particularly well secured, and so it remains an attractive payload to cybercriminals. Attackers exploit these vulnerabilities to utilize DNS as a platform for cyberattacks.

DNS-based cyberattacks strictly depend on Command-and-Control (C&C) servers to function. Attackers initially infuse command server IP addresses into their malware program which provides them complete control as well as an operational connection for malicious purposes. When cybercriminals use encrypted IP addresses security administrators can trace them since such addresses tend to appear suspicious. Attackers use DNS resolution methods to uncover C&C server IP addresses so malware can bypass security detections that utilize IP address blacklisting. Malware operational security includes its capability for executing Domain Generation Algorithms that generate many semi-random domain names to establish communication with servers discreetly.

People who monitor harmful domain names perform a vital role to secure Internet user safety while protecting sensitive information and preventing monetary and damage to public image losses. Research has shown that network traffic analysis along with C&C server communication interception are the traditional approaches to recognize dangerous domains. Web content analysis tools together with URL scanning have become standard techniques for identifying harmful activities. Researchers investigate the illumination of harmful domains through DNS analysis enhanced by default gateway addresses which serve to boost detection capabilities.

To provide a comprehensive overview of the approach used in previous studies, this article looks at it from two different perspectives.

[1-2-3] Shah Abdul Latif University Khairpur
[4] IBA – Institiue of Emerging Technologies Khairpur
[5] Benazir Bhutto Shaheed University of Technology & Skill Development Khairpur
[6] Sukkur IBA University
Country: Pakistan
Email: *mangisamar@gmail.com

This article discusses the strategies in two key elements in order to systematically present the methodologies utilized in earlier studies:

1. The experiment's data type to be used as the data source.
2. The technology that underlies the detecting technique.

The data sources used in earlier hostile domain identification techniques are described in Section 2 in detail. Four different categories of detection procedures were defined in Section 3. The article ends with an overview of the key conclusions and concepts.

DATA SOURCE

This section organizes the many data types utilized in the schemas suggested in the body of existing literature. Keep in mind that the type of data used is crucial in establishing the best malicious domain detection strategy.

A. DNS-related data

DNS-related data plays an important role in detecting and predicting malicious activity. This data can be divided into two categories, active DNS data and passive DNS data, depending on how it was collected.

Active DNS data is obtained by intentionally sending DNS queries and capturing the corresponding DNS responses. These packets are typically intercepted as they enter the network. Please note that requests are initiated by researchers, so this data may not accurately reflect actual user behavior. The main purpose of collecting active DNS data is to retrieve the DNS records for a particular domain. However, this data does not capture user behavior and is not suitable for malicious domain detection methods based on user-level capabilities. Additionally, active DNS data is often convenient to publish because it does not raise user privacy concerns. Researchers used Active DNS data for a variety of purposes. For instance, Khalil [4] conducts a thorough examination of live DNS data to pinpoint domains controlled by the same entity and create linkages between them. By observing DNS traffic, Segugio [5] built a binary graph known as the host domain graph to represent the behavior of host requests. By observing the DNS traffic produced by the Domain Generation Algorithm (DGA), Antonakakis et al. [6] suggested to extract a significant amount of NXDomain answer packets in cyberspace. We also extracted 33 features to parse and distinguish between DGA domain names and command and control (C&C) servers.

Overall, active DNS data provides valuable information and information for malicious domain detection and analysis, allowing researchers to discover patterns and correlations that help identify potential threats. Passive DNS data comes from actual DNS server records and captures historical mappings between domains and IP addresses. This type of data is more representative as it contains different characteristics and statistics. Unlike active DNS data, passive DNS data correlates with real user behavior and provides important user-level statistics to help identify malicious activity. One of the benefits of passive DNS data is that it provides information not available through active data collection methods. One limitation, however, is that potentially malicious domains can only be detected through passive links after registration. Previous research used negative DNS data to detect malicious DNS. Buildings et al. [7] were the first to use negative DNS data for this purpose, but their approach did not include domain reputation calculations. Notos [8] introduces dynamic reputation computation for unknown domains by extracting domain features. To address privacy concerns, Kopis [9] performs malicious domain detection by passively monitoring top-level DNS data rather than local DNS traffic. Bao et al. [10] used passive DNS data for analysis and developed a comprehensive feature analysis scheme based on domain name access and character properties.

External data sources are frequently used to leverage DNS data and serve as an important baseline for identifying fraudulent domains. These external sources include network data, related resource (RR) records, registration records, IP/domain/whitelist lists, geolocation data, and autonomous system numbers (ASNs). To create the dataset, researchers Wang [29] and his Jauniarovic et al. Combined. Improved. Additionally, our verification and training techniques for detecting fraudulent domains are based on trustworthy credibility data. This actual data may be considered an external source.

B. DGA related data

A Domain Name Generation Algorithm (DGA) uses cryptographic technology (such as MD5 or an XOR operation) to generate a series of pseudo-random strings that are valid lists of domain names. Hackers use this method to force malware to share the same random seed, giving them an approximate list of domain names. The malware then uses these domain names to establish connections with command-and-control servers until a connection is successfully established. Binary classification techniques are frequently used in research to distinguish between representative sets of algorithmically generated malicious and clean domains. Genuine internet services are associated with clean domains. To create a clean domain dataset, researchers often use popular lists provided by trusted Internet portals such as Alexa. Alexa offers different rankings sorted according to different criteria, including rankings by country. Obtaining a large number of negative samples of algorithmically generated domain names (harmless domain names) poses a problem for real-world researchers. To address this limitation, researchers often rely on publicly available datasets for their experiments. For example, the DGArchive [15] dataset is popular because it contains 62 different domain name strings generated algorithmically Selvi and colleagues [16] backed up their claims using the Bader repository, which provides thousands of DGA domain names from 26 different DGA implementations. The 360 netlab dataset and the Bambenek repository were complementary sources used by Ren et al. [17] Pooled DGA coverage. Similarly, Mao [18] used Network Lab 360 data containing 40 DGA series. These public datasets provide a valuable resource for researchers to study and analyze algorithmically-generated domain names, compensating for the difficulties of collecting large numbers of negative samples in real-world scenarios.

DETECTION METHOD

In this part, we thoroughly summarize the latest malware detection techniques and categorize them into four distinct groups: experimental, machine learning-based, deep learning-based, and graph-based.

A. Heuristic Approaches

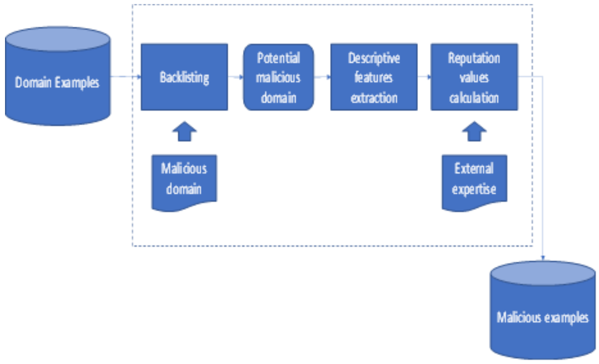
Domain Examples
Backlisting
Potential malicious domain
Descriptive features extraction
Reputation values calculation
Malicious domain
External expertise
Malicious examples

Figure 1: The structure of the heuristic approach contains the methods and principles used in these discovery schemes

To distinguish between malignant and benign areas, draw conclusions based on external experience. Researchers examined the raw data to find the most useful signatures that could pinpoint fraudulent sites. These techniques are sometimes called knowledge-based techniques [19]. Figure 1 shows the structure of the inference method.

According to McGrath, the domain names of legitimate and phishing sites are very different [20]. Identify phishing schemes using URL-related properties such as URL length and character syntax. Sandip [21] devised a technique to examine the distribution of uppercase and alphanumeric characters within domains connected to the same set of his IP addresses. A score is calculated based on this analysis to determine the quality of the field. However, using the domain name feature alone has limitations and may not be effective at detecting malicious domain names in general. User behavior often provides valuable information for determining a domain's reputation. Tan and Dong [22] created a simple approach called Domain Observers to use passive traffic data to examine access patterns between Internet users and domains. Blacklists [23] are also useful tools for early malware protection. To achieve efficient and accurate detection, Zhao et al. [24] introduced his two-step detection mechanism. Edit spaces are used for early and rapid filtering to reduce the addition time after a detected domain name is compared to a blacklist of known malicious URLs. Blacklist and whitelist filtering are frequently used in OMDD [25] to significantly reduce the amount of actual data.

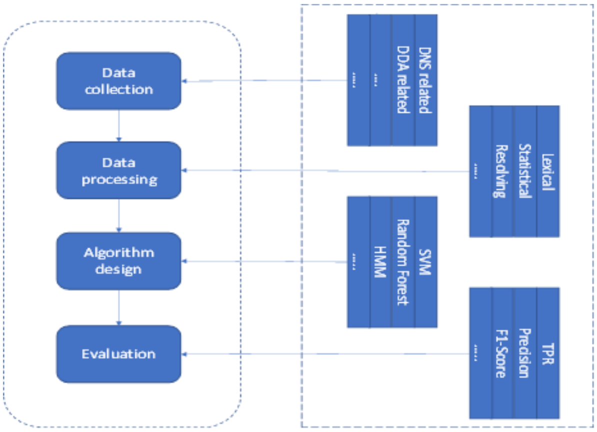

Figure 2: Structure of the machine learning approach. We describe the approach used in these detection scenarios that utilizes machine learning techniques to classify domains as malicious or safe

Most malicious domain detection methods are based on machine learning algorithms that can automatically distinguish between malicious and safe domains. supervised learning and unsupervised learning are the two basic categories into which these algorithms fall. Supervised learning is good when you have a large dataset with labeled examples, while unsupervised learning is useful when there are no labels available for a given dataset. In either case, feature engineering plays an important role in choosing suitable features for machine learning models. The machine learning-based method structure consists of four main phases as shown in the diagram. 2: Data collection, data processing, algorithm development and evaluation. During the data collection phase, relevant data sets are collected from publicly available sources or through specific data collection methods. Data preprocessing involves cleaning, transforming, and preparing data to make it suitable for machine learning algorithms. Algorithm development requires the selection and implementation of an appropriate machine learning model, given specific requirements and data characteristics. Finally, the performance and effectiveness of machine learning approaches in detecting malicious domains are evaluated. According to previous research, fixed lexical functions are well suited for simple machine learning-based techniques. Domain features have been grouped into his three categories by Kidmos et al. [26]: General characteristics, simple lexical characteristics, and high-level lexical adjectives. They used a random forest classifier and performed a 10-fold cross-validation on the model. This result demonstrates the effectiveness of the vocabulary feature in improving malicious domain detection performance. Galati [27] proposed a random forest classification model based on static lexical features for classifying malicious and safe URLs. URLs can be analyzed using three distinct types of features: blacklist-based features, lexical attributes, and host-related characteristics. Researchers have developed an improved version of the traditional machine learning classifier to overcome the limitations in detecting malicious domains, resulting in more robust and efficient performance. Zhu and Zou noted that standard Support Vector Machine (SVM) models tend to lose accuracy gradually over time. F-SVM represents their enhanced version that they used to resolve this issue. The new algorithm integrates feedback learning methodologies that improve accuracy while minimizing the expenditure of updating the model detection system. Researchers Tang and Dong utilized Baum-Welch and Viterbi algorithms to create a successful hidden Markov Model (BVHMM) platform which detected malicious domains. Researchers optimized an advanced Hidden Markov Model (HMM) to process large datasets efficiently. KSDom introduced an essential aspect in collaboration with the diagnosis framework in the detection process. The system obtains extensive DNS traffic data along with external DNS-related datasets. The KSDom system addresses data imbalance by combining k-means clustering with the Synthetic Minority Oversampling Technique (SMOTE) which enables CatBoost algorithm classification. The detection method delivers precise results for identifying harmful domains. Active DNS data shows strong potential to serve as a key threat detection element for creating standardized patterns which identify harmful domain names effectively.

C. Deep Learning Approach

Deep learning is a sophisticated approach that builds on traditional machine learning methods. It has achieved impressive results in areas like speech and image recognition, showcasing its promise for use in cybersecurity too. Additionally, deep learning is becoming a promising avenue for cancer diagnostics, opening up new opportunities for improving detection technologies. Deep His learning's approach can be modified for different application contexts using different mechanisms to improve the accuracy of malicious domain name recognition. Using these methods, researchers only need to provide a deep learning architecture containing raw domain names. These structures automatically and implicitly extract the most important and actionable properties. on the illustration. The structure of the deep learning approach is shown in Figure 3.

Vinayamar et al. [30] evaluated various deep learning techniques, including recurrent neural networks (RNN), long-term memory (LSTM), and traditional machine learning classifiers. They use local DNS records as a dataset for evaluation. Among deep learning techniques, LSTM shows the highest efficiency in detecting malicious domains. However, when dealing with very long LSTM domain names, it can be difficult to learn views efficiently. To address this limitation, another study presented an LSTM model and attentional mechanisms [31]. The attention mechanism allows the model to focus on more important subchains of the domain, thus improving representation and representation of the domain. Deep convolutional neural networks (CNNs) have shown excellent data processing efficiency with consistent transform properties. A variable autoencoder (VAE) was used by Sun et al. [32] used CNN to distinguish between malignant and benign domains to extract hidden information from features. To develop a new N-gram-based composite field classification (n-CBDC) model, Shaw et al. [33] integrated N-gram and CNN analysis. This model successfully detects pronounceable domain names based on the list of words generated by the Domain Creation Algorithm (DGA) and returns good results.

Harness the power of deep learning models to automatically extract and learn relevant features from input domain names. This framework eliminates the need for manual feature engineering because deep learning architectures inherently capture the most useful view. on the diagram. Figure 3 is a framework for a deep learning approach showing data flow and processing flow within the architecture.

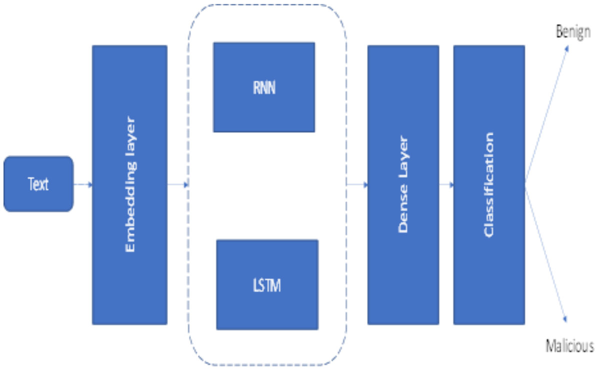

Figure 3: A framework of deep learning-based techniques covering common approaches and architectures used in these discovery schemes

D. Graphical Approach

Graph-based methods utilize ideas from network theory and graph theory to explore how individuals, groups, and domains interact. These techniques use domain IP links to calculate domain reputation scores and identify malicious domains, unlike machine learning classifiers that rely on supervised learning. Clients, domains, and IP addresses are the three different types of hosts that make up the domain IP address diagram in Figure 4. This graph-based approach offers an alternative to traditional network topology and link analysis for identifying hostile domains. A limitation of classification-based detection methods based on local domain name properties is that they are vulnerable to attackers who deliberately spoof DNS packets to avoid detection. However, using domain-IP relationships effectively overcomes this drawback. Graph theory offers path-based inference methods specifically designed to analyze DNS data and establish connections between domains by leveraging features derived from DNS datasets. Carrier et al. [4] carried out a comprehensive study of active DNS data to create reliable domain associations and enhance the effectiveness of empirical approaches for identifying malicious domains. Deep learning methods can also be effectively applied to graph-based analysis. DeepWalk, a prominent graph embedding algorithm, generates strong representations suitable for statistical modeling. Someone Wait [34] developed an insightful domain connectivity graph and utilized an enhanced version of DeepWalk to derive localized structural characteristics. The process of constructing graphs is a fundamental aspect of graph analysis. HinDom [11], a robust domain detection framework, addresses these challenges by building a heterogeneous information network (HIN) that integrates elements such as clients, domains, IP addresses, and their relationships, enabling a comprehensive analytical perspective. HGDom [12] incorporates a graph convolutional network (GCN) to manage the graph structure and node attributes in the HIN, while also leveraging a metapath-based attention mechanism to refine analysis.

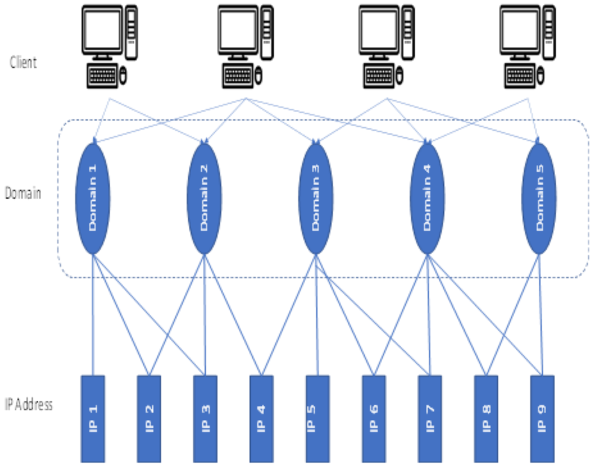

Figure 4: An example of an IP domain diagram showing the relationship between domains, IP addresses and clients

This diagram illustrates how different entities like clients, domains, and IP addresses are linked together. It provides a visual overview of the network's organization and the relationships within the IP ecosystem of the domain. The graphs depict how these elements are connected and how they relate to one another, which helps in understanding the interactions and behaviors of domains, particularly when it comes to identifying malicious domains.

CONCLUSION:

Malicious domains are a significant threat to cyber security and are often linked to various illegal activities. This article aims to provide a detailed look at previous research focused on detecting malicious domain names, examining two main areas: data resources and detection methods. When it comes to data resources, we can classify them into two categories: DNS-related data and data related to Domain Generation Algorithms (DGAs). DNS-related data is further divided based on how it's collected: active DNS and passive DNS. Active DNS data, while it doesn't reflect real user behavior, helps protect user privacy and is easily shared. On the other hand, passive DNS data comes from logs of DNS servers and provides comprehensive statistical insights into user behavior, which can be very helpful for detecting malicious domains. Data related to DGAs includes lists of domains associated with malicious activities, such as those involved in botnet operations. These datasets are often publicly available and are frequently used for analysis. However, each type of dataset has its own strengths and weaknesses, making it challenging to use the available data to improve detection accuracy and efficiency. In terms of detection methods, we categorize them into four main types: machine learning, deep learning, heuristic, and graph-based methods. Heuristic methods rely on expert knowledge to identify potentially harmful domains, but they can be time-consuming and may lack precision. Machine learning techniques that focus on extracting features have shown promise in identifying malicious domains, but they often fall short in real-time detection and response. Deep learning methods, however, offer a more advanced approach by allowing raw domain names to be processed directly within sophisticated learning frameworks. This results in higher accuracy and better performance. Finally, graph-based methods analyze the relationships between domains, calculate reputation scores, and effectively identify domains with malicious intent. The creation of meaningful graphs is essential for successful graph analysis. Future research should aim to develop efficient real-time detection systems and tackle the issue of data imbalance. Additionally, exploring effective combinations of existing detection systems and creating innovative methods could lead to improved detection outcomes. Through continued research, we hope to enhance efforts against illegal cyber activities linked to malicious domains.

REFERENCE

[1] Rice, T., Kim, D. W., & Yang, M. (2023, April). Developing a GUI Application: GPU-Accelerated Malicious Domain Detection. In Proceedings of the 2023 ACM Southeast Conference (pp. 167-171).

[2] Wagan, A. A., Li, Q., Zaland, Z., Marjan, S., Bozdar, D. K., Hussain, A., ... & Baryalai, M. (2023). A Unified Learning Approach for Malicious Domain Name Detection. Axioms, 12(5), 458.

[3] Luo, H., Liu, W., & Cao, Q. (2023, June). DGA domain name detection model based on multiscale feature. In Second International Symposium on Computer Applications and Information Systems (ISCAIS 2023) (Vol. 12721, pp. 53-61). SPIE.

[4] Khalil, I. M., Guan, B., Nabeel, M., & Yu, T. (2018, March). A Domain is only as Good as its Buddies. In Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy. ACM.

[5] Rahbarinia, B., Perdisci, R., & Antonakakis, M. (2015, June). Segugio: Efficient behavior-based tracking of malware-control domains in large ISP networks. In 2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (pp. 403-414). IEEE.

[6] Antonakakis, M., Perdisci, R., Nadji, Y., Vasiloglou, N., Abu-Nimeh, S., Lee, W., & Dagon, D. (2012). From {Throw-Away} traffic to bots: Detecting the rise of {DGA-Based} malware. In 21st USENIX Security Symposium (USENIX Security 12) (pp. 491-506).

[7] Zdrnja, B., Brownlee, N., & Wessels, D. (2007, July). Passive monitoring of DNS anomalies. In International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (pp. 129-139). Berlin, Heidelberg: Springer Berlin Heidelberg.

[8]. Antonakakis, M., Perdisci, R., Dagon, D., Lee, W., & Feamster, N. (2010). Building a dynamic reputation system for {DNS}. In 19th USENIX Security Symposium (USENIX Security 10).

[9]. Antonakakis, M., Perdisci, R., Lee, W., Vasiloglou II, N., & Dagon, D. (2011). Detecting malware domains at the upper {DNS} hierarchy. In 20th USENIX Security Symposium (USENIX Security 11).

[10]. Bao, Z., Wang, W., & Lan, Y. (2019, August). Using passive dns to detect malicious domain name. In Proceedings of the 3rd International Conference on Vision, Image and Signal Processing (pp. 1-8).

[11]. Sun, X., Tong, M., Yang, J., Xinran, L., & Heng, L. (2019). {HinDom}: A robust malicious domain detection system based on heterogeneous information network with transductive classification. In 22nd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2019) (pp. 399-412).

[12]. Sun, X., Yang, J., Wang, Z., & Liu, H. (2020, April). HGDom: heterogeneous graph convolutional networks for malicious domain detection. In NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium (pp. 1-9). IEEE.

[13]. Stone-Gross, B., Cova, M., Cavallaro, L., Gilbert, B., Szydlowski, M., Kemmerer, R., ... & Vigna, G. (2009, November). Your botnet is my botnet: analysis of a botnet takeover. In Proceedings of the 16th ACM conference on Computer and communications security (pp. 635-647).

[14]. Alexa. https://www.alexa.com

[15]. Plohmann, D., Yakdan, K., Klatt, M., Bader, J., & Gerhards-Padilla, E. (2016). A comprehensive measurement study of domain generating malware. In 25th USENIX Security Symposium (USENIX Security 16) (pp. 263-278).

[16]. Selvi, J., Rodríguez, R. J., & Soria-Olivas, E. (2019). Detection of algorithmically generated malicious domain names using masked N-grams. Expert systems with applications, 124, 156-163.

[17] Ren, F., Jiang, Z., Wang, X., & Liu, J. (2020). A DGA domain names detection modeling method based on integrating an attention mechanism and deep neural network. Cybersecurity, 3(1), 4.

[18] Mao, J., Zhang, J., Tang, Z., & Gu, Z. (2020). DNS anti-attack machine learning model for DGA domain name detection. Physical Communication, 40, 101069.

[19] Zhauniarovich, Y., Khalil, I., Yu, T., & Dacier, M. (2018). A survey on malicious domains detection through DNS data analysis. ACM Computing Surveys (CSUR), 51(4), 1-36.

[20] McGrath, D. K., & Gupta, M. (2008). Behind Phishing: An Examination of Phisher Modi Operandi. LEET, 8, 4.

[21] Yadav, S., Reddy, A. K. K., Reddy, A. N., & Ranjan, S. (2010, November). Detecting algorithmically generated malicious domain names. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement (pp. 48-61).

[22] Tang, H., & Dong, C. (2019). Detection of malicious domain names based on an improved hidden Markov model. International Journal of Wireless and Mobile Computing, 16(1), 58-65.

[23] Kührer, M., Rossow, C., & Holz, T. (2014). Paint it black: Evaluating the effectiveness of malware blacklists. In Research in Attacks, Intrusions and Defenses: 17th International Symposium, RAID 2014, Gothenburg, Sweden, September 17-19, 2014. Proceedings 17 (pp. 1-21). Springer International Publishing.

[24] Zhao, H., Chang, Z., Wang, W., & Zeng, X. (2019). Malicious domain names detection algorithm based on lexical analysis and feature quantification. IEEE Access, 7, 128990-128999.

[25] Cui, J., Zhang, L., Liu, Z., Li, J., & Shi, L. (2018, October). An efficient framework for online malicious domain detection. In 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (pp. 1-6). IEEE.

[26] Kidmose, E., Stevanovic, M., & Pedersen, J. M. (2018, June). Detection of malicious domains through lexical analysis. In 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security) (pp. 1-5). IEEE.

[27] Ghalati, N. F., Ghalaty, N. F., & Barata, J. (2020). Towards the detection of malicious URL and domain names using machine learning. In Technological Innovation for Life Improvement: 11th IFIP WG 5.5/SOCOLNET Advanced Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2020, Costa de Caparica, Portugal, July 1–3, 2020, Proceedings 11 (pp. 109-117). Springer International Publishing.

[28] Zhu, J., & Zou, F. (2019, August). Detecting malicious domains using modified SVM model. In 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 492-499). IEEE.

[29] Wang, Q., Li, L., Jiang, B., Lu, Z., Liu, J., & Jian, S. (2020). Malicious domain detection based on k-means and smote. In Computational Science–ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part II 20 (pp. 468-481). Springer International Publishing.

[30] Vinayakumar, R., Soman, K. P., & Poornachandran, P. (2018). Detecting malicious domain names using deep learning approaches at scale. Journal of Intelligent & Fuzzy Systems, 34(3), 1355-1367.

[31] Chen, Y., Zhang, S., Liu, J., & Li, B. (2018, September). Towards a deep learning approach for detecting malicious domains. In 2018 IEEE International Conference on Smart Cloud (SmartCloud) (pp. 190-195). IEEE.

[32] Sun, Y., Chong, N. S., & Ochiai, H. (2020, March). Text-based malicious domain names detection based on variational autoencoder and supervised learning. In 2020 54th Annual Conference on Information Sciences and Systems (CISS) (pp. 1-5). IEEE.

[33] Xu, C., Shen, J., & Du, X. (2019). Detection method of domain names generated by DGAs based on semantic representation and deep neural network. Computers & Security, 85, 77-88.

[34] He, W., Gou, G., Kang, C., Liu, C., Li, Z., & Xiong, G. (2019, October). Malicious domain detection via domain relationship and graph models. In 2019 IEEE 38th International Performance Computing and Communications Conference (IPCCC) (pp. 1-8). IEEE.

[35] Cersosimo, M., & Lara, A. (2022, April). Detecting malicious domains using the splunk machine learning toolkit. In NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium (pp. 1-6). IEEE.

[36] Park, K. H., Song, H. M., Do Yoo, J., Hong, S. Y., Cho, B., Kim, K., & Kim, H. K. (2022). Unsupervised malicious domain detection with less labeling effort. Computers & Security, 116, 102662.

[37] Atrees, M., Ahmad, A., & Alghanim, F. (2022). Enhancing Detection of Malicious URLs Using Boosting and Lexical Features. Intelligent Automation & Soft Computing, 31(3).

[38] Darwish, S. M., Anber, A. E., & Mesbah, S. (2021). Bio-inspired machine learning mechanism for detecting malicious URL through passive DNS in big data platform. Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges, 147-161.

[39] Silveira, M. R., da Silva, L. M., Cansian, A. M., & Kobayashi, H. K. (2021, December). Detection of newly registered malicious domains through passive DNS. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 3360-3369). Ieee.

[40] Deshpande, A., Pedamkar, O., Chaudhary, N., & Borde, S. (2021). Detection of phishing websites using Machine Learning. International Journal of Engineering Research & Technology (IJERT), 10(05).