Web Phishing Detection System Based on Artificial Neural **Networks Technology**

Vitaliy Tsyganok^{1,2}, Yaroslav Khrolenko¹, Iryna Domanetska², Olena Fedusenko² and

Julia Minaeva²

¹ Institute for Information Recording of NAS of Ukraine, 2 Shpaka Street, Kyiv, 03113, Ukraine ² Taras Shevchenko National University of Kyiv, 24 Bogdana Gavrylishina Str,04116, Kyiv, Ukraine

Abstract

The widespread use of Internet technologies, in addition to the general positive in the context of the development of society, has led to the emergence and rapid growth of criminal activities carried out with the use of high technologies. Currently, phishing is one of the most common types of Internet crime. The task of detecting phishing is urgent, because phishing attacks lead to large losses due to the malicious use of personal data, confidential information, commercial or state secrets. The paper examines modern threats and methods of countering phishing attacks, analyzes available methods and means of protection, and proposes a method of protecting against phishing attacks using neural networks. The uniqueness of the approach proposed in the article to solving the problem of detecting phishing links lies in the use of hybrid architectures of neural networks, namely a combination of convolutional and recurrent neural networks. The resulting hybrid architecture demonstrates a higher indicative accuracy compared mono-architecture convolutional and recurrent neural networks.

Keywords¹

phishing, artificial neural network, convolutional neural network, recurrent neural network, conditional random fields, long-short-term memory network, hybrid neural network.

1. Introduction

The accelerated growth of information technologies all around the world and in Ukraine, especially observed in the last decade, is inevitably accompanied by the dynamic development of crimes in this sphere. Along with global computerization and the development of digital technologies, which greatly simplified human life, the concept of cybercrime has entered our lives. Cybercrimes are the most dynamic group of socially dangerous acts because every year cybercrimes become more and more widespread and dangerous. Today, almost all experts in the field of information technology acknowledge that the situation with cybercrime in the world is getting worse.

Phishing remains the most massive threat to Ukrainian Internet users, and its scale is growing. Notably, phishing sites account for 88% of blocked resources, while the remaining 12% are fraudulent online stores, fraudulent money-making schemes, "investment" and service fraud that extort money from citizens, and sites with malicious software. Nowadays, this problem is becoming even more urgent in Ukraine. During 2022, the number of Russian phishing attempts against Ukraine increased by 250%. The main target of Russian hackers was more than 150 government institutions, with the Ministry of Defense of Ukraine being the primary target. Therefore, the creation of effective software tools for detecting phishing links is an urgent problem.

International Scientific Symposium "Intelligent Solutions" (IntSol-2023), September 27-28, 2023, Kviv-Uzhhorod, Ukraine

EMAIL: vitaliy.tsyganok@gmail.com (A.1), domanetska@knu.ua (A.2),yaroskhr@gmail.com (A.3); fedusenko@knu.ua (A.4), julia.minaeva@knu.ua (A.5)

ORCID: 0000-0002-0821-4877 (A.1); 0000-0002-8629-9933 (A.1); 0009-0004-0641-827X (A.2); 0000-0002-5782-5922 (A.3); 0000-0002-2367-1507 (A.4)

 ²⁰²³ Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

2. The aim of the study

Analysis of the distinctive features of the process of protection against phishing attacks, along with a comprehensive review of existing approaches to solving the problem of identification of phishing web links using computational intelligence tools, development of a software application based on a classifier of Internet links based on a hybrid artificial neural network.

3. Materials and research methods

The research is focused on seeking an effective combination of two different neural network architectures to solve the task of classifying internet links for phishing link identification. The study involved an analysis of mono-architectures of convolutional networks (CNN) and recurrent networks (SimpleRNN), hybrid models, namely combinations of convolutional and recurrent networks, deep forward propagation networks (DNN), and recurrent neural networks based on LSTM elements. The dataset used to evaluate the performance of these hybrid architectures for phishing link identification was obtained from the UCI Machine Learning Repository and comprised approximately 2500 labeled examples. This dataset is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

The project was implemented using Python programming language and the Keras framework.

4. Literature review

According to the Law of Ukraine "On the Fundamental Principles of Cybersecurity of Ukraine," cybersecurity is defined as the protection of the vital interests of individuals, citizens, society, and the state during the use of cyberspace. It ensures the sustainable development of the information society and digital communication environment, and the timely detection, prevention, and neutralization of real and potential threats to Ukraine's national security in cyberspace.

A cyber threat encompasses a combination of factors and conditions that pose risks to information security. Malicious actors may target IT infrastructure, workstations, mobile devices, other technical tools, and ultimately, individuals as elements in cyberspace. Phishing is an automated form of social engineering used by malefactors to exploit the Internet to deceptively acquire confidential information from companies and individuals, often impersonating legitimate websites [1].

Malicious actors send harmful email attachments or URLs to users, seeking access to their accounts or computers [2]. Cybercriminals have become very sophisticated, with many emails escaping spam detection. Users receive emails, allegedly requiring them to change passwords or update payment information, unintentionally granting criminals access to confidential data.

The high potential for gains, such as accessing bank accounts and credit card numbers, the ease of disseminating forged emails posing as legitimate authorities, and the challenges faced by law enforcement in apprehending such criminals, have led to a surge in phishing attacks in recent years.

The "State of the Phish" report [3] for the year 2019 revealed that nearly 90% of organizations experienced targeted phishing attacks during that year. 84% reported phishing through SMS/text, 83% encountered voice phishing, and the email phishing volume grew by 67% in a year. These data indicate a rising trend of people avoiding internet commerce due to identity fraud concerns, despite companies taking on the risk of fraud. According to Microsoft's annual report, the number of cyberattacks increased by 3.5 times in 2022 compared to 2021. Financial institutions, social media platforms, payment systems, and e-commerce are the most attractive targets for phishing (Figure 1).

Researchers discuss and propose a variety of solutions to overcome phishing challenges, yet there is no solution that can be trusted or used for fully mitigating these attacks. The anti-phishing measures, proposed in the literature can be categorized into three main defense strategies.

The first line of defense assumes human-factor solutions that educate end-users to recognize phishing attempts and avoid falling victim to them.

The second line of defense comprises technical solutions developed to avert attacks at early stages, such as vulnerability levels, to prevent threats from materializing on user devices, by reducing human

impact and detecting attacks. This also involves employing specific methods to detect the source of attacks (e.g., identifying newly registered domains closely resembling well-known domain names).

The third line of defense assumes the involvement of law enforcement agencies as a restraining control. These approaches can be combined to create significantly stronger anti-phishing solutions.

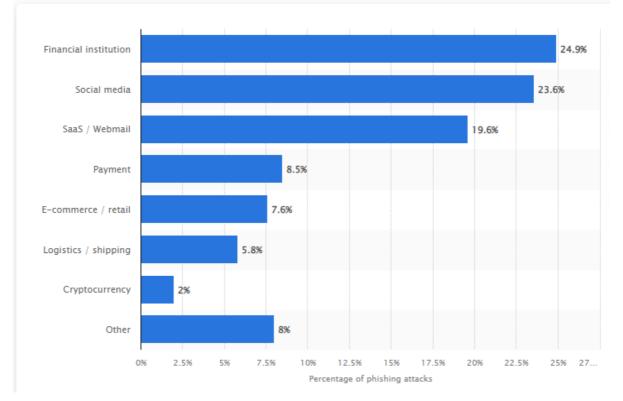


Figure 1: Distribution of phishing attacks by targets [4]

Human education is an efficient countermeasure to elude and avert phishing attacks. Awareness and education are the first lines of defense in anti-phishing methodology, even if it does not provide complete elimination of the threats. End-user training reduces users' susceptibility to phishing attacks and complements other technical solutions. According to the analysis conducted [5], 95% of phishing attacks are caused by the human factor. There are various technical solutions for eliminating phishing threats. The proposed technical solutions for detecting and stopping phishing attacks can be presented by two main approaches: content-aware solutions and content-based solutions. Content-aware methods include blacklists and whitelists which classify faked emails or web pages based on information which is not a part of the email or web page, such as URLs and domain name features. The disadvantage of this approach is the impossibility to identify all phishing websites because after deleting the phishing site, the phisher can easily register a new domain [6, 7].

Content-based methods classify a page or email according to the information in its content, for instance, text, images, HTML, JavaScript, or cascading style sheets (CSS). Content-based solutions incorporate machine learning, heuristics, visual similarity, and image processing techniques [8].

Finally, multi-aspect methods use a combination of the previous approaches to detect and avert phishing attacks. Lately, there has been a tendency of applying ML technologies in the implementation of anti-phishing solutions for the early detection of phishing threats and minimizing the risks of danger. Currently, security strategies based on neural network technologies are becoming more and more widespread [9-11]. The article [12] reviewed 16 classification systems based on the semantic characteristics of URLs. Ten characteristics that distinguish safe websites from phishing websites were also collected and analyzed using semantic features. According to the results of the comparison, GradientBoostingClassifier and RandomForestClassifier showed the highest accuracy. The researchers noted that one possible limitation is the task of feature selection. The study [13] treats malicious URL detection as a binary classification problem and examines the performance of known

classifiers (Naive Bayes, Support Vector Machines, Multilayer Perceptron, Decision Trees, Random Forest, and k-Nearest Neighbors).

The authors of [14] focused on semantic feature extraction methods using word2vec to improve the description of the features of phishing sites, and then combined these features with other statistical features to create a more robust phishing detection model. Experimental results with actual datasets have shown that feature combinations improve phishing detection performance.

The authors of [15] compare the random forest method and recurrent neural networks within the task URL classifications. Neural networks have shown better efficiency, so the authors came to conclusion about preference.

Separately, it should be said about teaching without a teacher: despite the low accuracy, this approach is now one of the promising directions of scientific research. Such models do not require tagged data and therefore have great potential for application, including in the field of cyber security. It is obvious that the accuracy of the classifier is not sufficient for independent use in real conditions. However, given the fact that the data was not labeled, there is no need to train the model, which is a good result. Similar classifiers may be required as an additional mechanism within the framework of a heuristic approach to identify suspicious links that should be paid attention to and possibly subjected to a deeper investigation. It is also possible to resort to clustering to help mark objects, which will later, after being checked by an expert, be used for training the classifier model (training with a teacher) or for deterministic methods — the search for an unambiguous match with the detected phishing link. Criminals are constantly changing their attack tactics to exploit system vulnerabilities and user ignorance. Choosing an inappropriate countermeasure algorithm can lead to unpredictable results and wasted effort, which ultimately affects the accuracy and effectiveness of the Deep Cybersecurity DL model [16].

This is another argument in favor of the use of deep learning algorithms, because DL algorithms provide the possibility of automating the process of detecting signs of phishing links, provide flexibility and adaptability to changes, of course, the need for additional retraining of the algorithm.

Nowadays, there are many different DL algorithms that numerous researchers have implemented to detect phishing websites. There are various DL-based approaches designed to solve a specific problem or meet certain system requirements; each has its advantages and disadvantages [17, 18]. Although deep learning techniques, and artificial networks in particular, take a long time to train, they often provide greater accuracy and automatically extract features from raw data without any prior knowledge. However, selecting the right approach that is best suited for a certain application or data set is a challenging task. Different performance measures were used to evaluate the effectiveness of the DL-based phishing detection model. The indicators obtained as a result of the experiments indicate that among the four DL methods (DNN, CNN, LSTM, GRU) no algorithm gave the best values for all performance indicators. One should choose the one that best suits their specific applications or according to their specific requirements.

In this research, the possibility of combining different neural network architectures into a hybrid or ensemble model was investigated to achieve the advantages and eliminate the disadvantages of monoarchitectural artificial neural networks. Based on the conducted research, we can conclude that a promising direction in the task of increasing the effectiveness of phishing detection is the use of hybrid models, in particular, the models that combine layers of different nature [4].

Promising combinations for research might encompass the following:

- 1) CNN + DNN
- 2) DNN +LSTM
- 3) CNN +RNN

The dataset from [19] was used for the experiments. The augmented dataset consists of 10,000 instances obtained from 5,000 phishing and 5,000 legitimate websites.

eneral description of attributes:

- having IPAddress checking for the availability of an IP address in the link;
- URLLength checking the number of characters in the link;
- ShorteningService checking if the link is displayed in a shortened format;
- havingAtSymbol checking if the link has the sign "@";
- doubleslashredirecting checking if the link contains the sign "\\";
- PrefixSuffix checking if there is no prefix or suffix attached to the link;

- havingSubDomain checking if no subdomain is attached to the link;
- SSLfinalState verification of the SSL certificate;
- Domainregistrationlength domain lifecycle check;
- Favicon check if its personal Favicon is attached to the link;
- Port checking if no ports are attached to the link and their status;
- HTTPStoken HTTPS certificate verification;
- RequestURL checking if no automatically downloaded data is attached to the link;
- URLofAnchor checking if "anchor" links are not attached;
- Linksintags checking if no SQL injections are attached to the link;
- SFH (server form handler) checking if no SFH injections are attached to the link;
- Submittingtoemai check for attachment to mail;
- AbnormalURL checking for a fake domain;
- Redirectpage checking for redirection to another page;
- onMouseOver checking for a hidden link;
- RightClick checking if the link is displayed as an "<a>" element;
- Using pop-upwidnow checking for pop-up windows;
- Iframe checking for the presence of an Iframe element;
- Ageofdomain checking for the length of the life cycle;
- DNSRecord checking for redirection through additional DNS servers;
- Webtraffic traffic volume checking;
- PageRank checking the link rating in the Black/White lists;
- GoogleIndex checking the link ranking in Google;
- Linkspointingtopage checking for the presence of "magnetic" links;
- Statisticalreport obtaining confirmation of security from open databases;

Python was selected as the main software platform for the development of the phishing link recognition system due to its effective collection of scientific tools.

Libraries used in the process of the system development included:

- numpy an extension of the Python language that adds support for large multidimensional arrays and matrices;
- Keras a high-level neural networks API that operates using such software tools for creating deep networks;
- matplotlib an extensive library for creating 2D visualizations;]
- pandas a Python library used for data manipulation and analysis;.

For the comparison of the efficiency of their work, the following scheme of the experiment was proposed:

- Stage 1 comparative analysis of networks built on a single (mono) basic model, namely CNN, RNN.
- Stage 2 comparative analysis of the winner network of the 1st stage with the CNN-RNN hybrid network.
- Stage 3 comparative analysis of the winner network of the 2nd stage with the DNN-LSTM hybrid network.

During the experiment, all the mentioned models were trained on the same input data and with the same number of epochs (30). At the first stage of the experiment, the efficiency of the convolutional and recurrent network was analyzed. The network architectures that participated in the experiment are presented in Figure 2 and Figure 3, respectively.

A detailed description of convolutional network architecture using Keras framework tools:

- convolution layer (Conv1D) size (batch_size) 200, filters 200;
- sampling layer (MaxPooling1D) size 2;
- selection layer of 20% of existing neurons (Dropout);
- vector reconstruction layer (Flatten);
- layer of neurons (Dense) size 2 with Relu activation function;
- layer of neurons (Dense) of size 1 with sigmoid activation function.

The approach is based on the work of a convolutional neural network at the symbol level. In particular, URL and DNS strings are converted to a vector form using natural language processing techniques. CNN is utilized for the extraction of phishing features and training a binary classification

model.A competitor of the convolutional network in the first experiment was a recurrent neural network (RNN). The choice was due to the fact that recurrent neural networks specialize in processing sequential data and are widely used for text processing. Input text is usually abstracted to a sequence of characters, words, or phrases. In our experiment - to symbols.

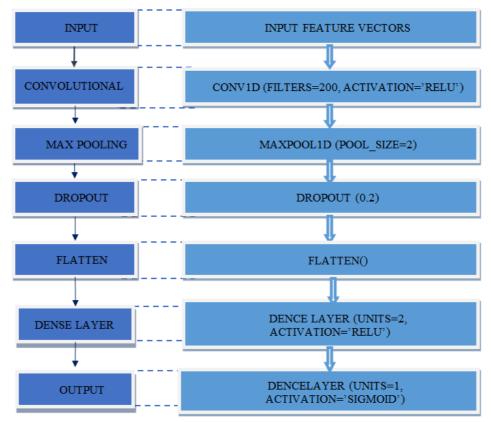


Figure 2: CNN architecture for phishing detection

A detailed description of the recurrent network architecture using Keras framework tools:

- recurrent layer (SimpleRNN) of 128 neurons, (batch_size) 200;
- a selection layer of 20% of existing neurons (Dropout);
- layer of neurons (Dense) size 2; a layer of neurons (Dense) of size 2 with the ReLU activation function;
- layer of neurons (Dense) of size 1 with sigmoid activation function.

The chosen SimpleRNN architecture has a basic form of RNN architecture. Contrary to the classic architecture proposed in many articles, the implementation of this model in Keras is completely different, but simple. Each RNN cell accepts one data input and has one hidden state that is passed from one step to the next. The results of comparing the effectiveness of monoarchitectures CNN and RNN (SimpleRNN) are presented in the **Figure 4**. Contrary to expectations, the CNN model showed the best result. Most likely, this is due to the simplicity of the recurrent network architecture used.

At the second stage, the monoarchitecture of the CNN (Figure 2) from the previous stage and the convolutional hybrid network were compared, in which the convolutional network is reinforced by a recurrent layer, the layers are connected in series (Figure 5).

Detailed description of the architecture of the hybrid CNN-RNN network:

- convolutional layer (Conv1D) size (batch_size) 200, filters (filters) 150;
- sampling layer (MaxPooling1D) size 2;
- recurrent layer (SimpleRNN) with 50 neurons;
- selection layer of 10% of existing neurons (Dropout);
- dense layer of size 1 with sigmoid activation function.

The effectiveness of the CNN and CNN-RNN models was evaluated by comparing the values of the accuracy metric. The best results were shown by the hybrid model (Figure 6). The next stage of the experiment included a comparison of CNN-RNN and DNN-LSTM hybrid architectures. Images of

neural network architectures are shown in Figure 5 and Figure 7, respectively. In this experiment, a complex ensemble architecture proposed by the authors [20] was used. This complex neural network consists of two parallel networks: a deep forward propagation network (DNN) and a recurrent network on LSTM blocks.

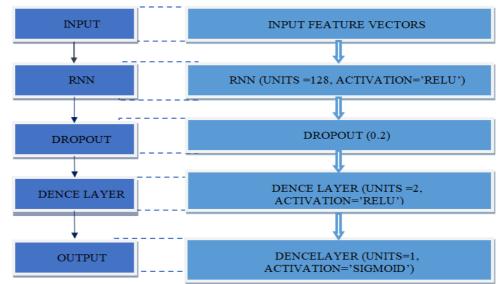


Figure 3: RNN architecture for phishing detection

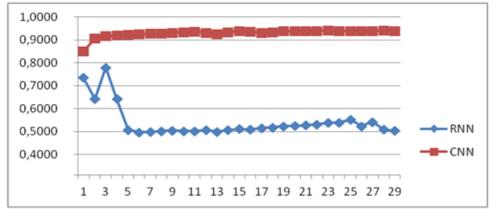


Figure 4: The effectiveness of CNN and RNN architectures by accuracy metric

The DNN architecture consists of four fully connected layers and a selection layer:

- dense layer of neurons of size 40 with ReLU activation function;
- dense layer of neurons of size 64 with ReLU activation function;
- dense layer of neurons of size 32 with ReLU activation function;
- selection layer of 20% of existing neurons (Dropout);
- dense layer of neurons of size 16 with ReLU activation function.

The LSTM network consists of two LSTM layers of 32 neurons each, a Dropout selection layer and a fully connected layer of 16 neurons. Both networks are connected by a fully connected layer of 8 nodes. The output Dense layer consists of 1 neuron with a sigmoidal activation function and calculates the result of the entire system (Figure 7). At the third stage, when comparing hybrid architectures (CNN-RNN and DNN-LSTM), the CNN-RNN model (Figure 8) showed the best results according to the accuracy metric, although the spread of values is very small.

5. Results

Experimental studies of solving the problem of identifying phishing Internet links on the "Phishing Websites Dataset" dataset from the UCI Machine Learning Repository have shown that simple mono-

architectural networks lose to hybrid ones. The table summarizes the results of the conducted experiments on the accuracy parameter on the test set (Table 1).

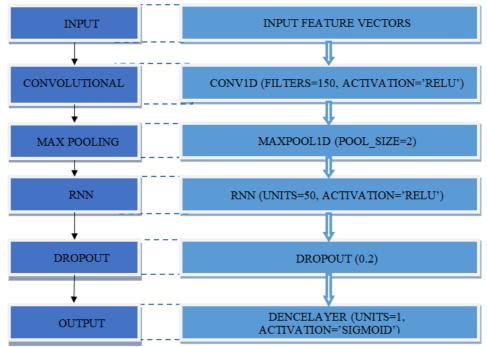
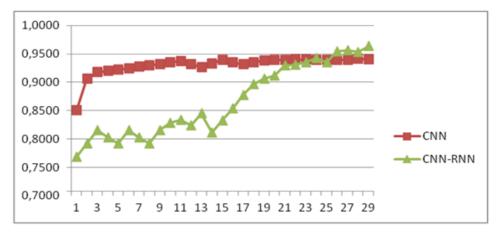
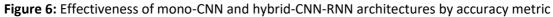


Figure 5: Hybrid CNN-RNN architecture for phishing detection





Among complex hybrid architectures, the CNN-RNN model is one of the most effective. When dealing with one-dimensional sequence data, CNN is extremely successful at extracting and achieving features. In a hybrid model with a sequential connection of convolutional and recurrent networks, the CNN interprets the input data of a sequence of symbols, extracts features from the input data, which are then sequentially transmitted to the RNN model for further understanding and classification. This combination of models provides a high level of flexibility and efficiency of the model.

Table 1

The effectiveness of different studied ANN-architectures

ANN-architectures	Accuracy
CNN	92%
RNN	50%
CNN + DNN	77%
DNN +LSTM	93%
CNN +RNN	97%

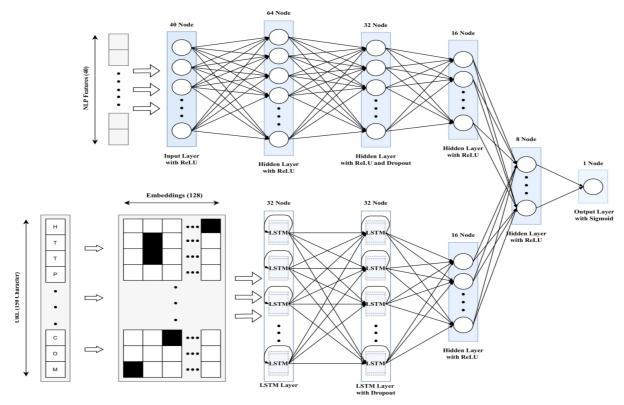


Figure 7: DNN-LSTM cture for phishing detection [20]

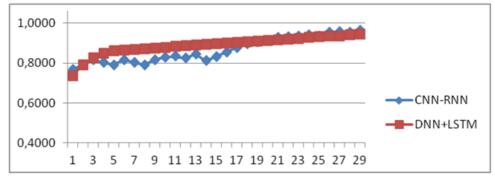


Figure 8: Effectiveness of CNN-RNN and DNN-LSTM architectures by accuracy metric

6. Conclusions

In the study, an analysis of existing research on the detection of phishing links based on the use of different types of neural network architectures was carried out. Architectures: CNN, DNN, RNN, LSTM networks were considered, their disadvantages and advantages were studied.

Training and testing of mono and hybrid models of neural networks was carried out. The following architectural models were compared according to efficiency indicators: CNN, RNN, CNN-RNN, DNN-LSTM. The computational experiment showed that the most effective model is the CNN-RNN network. The created neural network core is the basis of a software product for automating the detection and blocking of phishing links. Architecturally, the phishing link detection system is implemented as a browser extension. A promising area of future research is the development of neural network architectures using ensemble methods.

7. References

[1] Types of Cybercrime. Panda Security, 2023. URL: https://www.pandasecurity.com/en/mediacenter/panda-security/types-of-cybercrime/ (date of access: 24.06.2023)

- [2] Khrolenko V., Khrolenko Ya. Determination of phishing links using neural networks// 19 international scientific-practical conference «Management of the development of technologies» Kyiv, 28 March 2022; pp. 57 -58.
- [3] Egan, G. State of the Phish. Proofpoint. 2020. URL: <u>https://www.proofpoint.com/us/security-awareness/post/2020-state-phish-security-awareness-training-email-reporting-more-critical</u> (date of access: 24.06.2023)
- [4] . Khrolenko Ya. O. "Phishing detection system based on the use of neural networks" 2022. Thesis for obtaining a master's degree under the educational and professional program "Cybersecurity", Taras Shevchenko KNU, 2022 - 43 p.
- [5] Chen, Z. Deep Learning for Cybersecurity: A Review. In Proceedings of the 2020 International Conference on Computing and Data Science (CDS), Stanford, CA, USA, 1–2 August 2020; pp. 7–18. DOI:10.1109/CDS49703.2020.00009.
- [6] Siti Hawa Apandi, Jamaludin Sallim and Roslina Mohd Sidek. Types of anti-phishing solutions for phishing attack // IOP Conference Series: Materials Science and Engineering, Volume 769, The 6th International Conference on Software Engineering & Computer Systems 25-27 September 2019, Pahang, Malaysia. DOI:10.1088/1757-899X/769/1/012072
- [7] Said Baadel & Joan Lu, 2019. "Data Analytics: Intelligent Anti-Phishing Techniques Based on Machine Learning," Journal of Information & Knowledge Management (JIKM), World Scientific Publishing Co. Pte. Ltd., vol. 18(01), pages 1-17, March. DOI: 10.1142/S0219649219500059
- [8] Chanti, S., and Chithralekha, T. (2020). Classification of anti-phishing solutions. SN Comput. Sci. 1, 11. doi:10.1007/s42979-019-0011-2
- [9] Feng, J.; Zou, L.; Nan, T. A Phishing Webpage Detection Method Based on Stacked Autoencoder and Correlation Coefficients. Journal of Computing and Information Technology, Vol. 27, No. 2, June 2019, 41–54. DOI: 10.20532/cit.2019.1004702
- [10]Feng, Jian & Zou, Lianyang & Ye, Ou & Han, Jingzhou. (2020). Web2Vec: Phishing Webpage Detection Method Based on Multidimensional Features Driven by Deep Learning. IEEE Access. 8. 221214-221224. DOI: 10.1109/ACCESS.2020.3043188.
- [11]Huang, Y.; Yang, Q.; Qin, J.; Wen, W. Phishing URL Detection via CNN and Attention-Based Hierarchical RNN. In Proceedings of the 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science And Engineering (TrustCom/BigDataSE), Rotorua, New Zealand, 5–8 August 2019; pp. 112– 119. DOI: 10.1109/TrustCom/BigDataSE.2019.00024
- [12] Zhang, Xi & Zeng, Yu & Jin, Xiao-Bo & Yan, Zhi-Wei & Geng, Guang-Gang. (2017). Boosting the phishing detection performance by semantic analysis. DOI:1063-1070. 10.1109/BigData.2017.8258030.
- [13] Vanhoenshoven, F., Napoles, G., Falcon, R., Vanhoof, K., Koppen, M.: Detecting malicious URLs using machine learning techniques. In 2016 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–8, DOI: 10.1109/SSCI.2016.7850079.
- [14]Aksu, Doğukan & Turgut, Zeynep & Ustebay, Serpil & Aydin, M.Ali. (2019). Phishing Analysis of Websites Using Classification Techniques. DOI:10.1007/978-981-13-0408-8_21.
- [15]Zhao, Jingling & Wang, Nan & Ma, Qian & Cheng, Zishuai. (2019). Classifying Malicious URLs Using Gated Recurrent Neural Networks. DOI: 10.1007/978-3-319-93554-6_36.
- [16]Wei, Dabao & Feng, Jun. (2020). Network Attacks Detection Methods Based on Deep Learning Techniques: A Survey. Security and Communication Networks. 2020(1). 1-17. DOI: 10.1155/2020/8872923
- [17]Apruzzese, Giovanni & Colajanni, Michele & Ferretti, Luca & Guido, Alessandro & Marchetti, Mirco. (2018). On the effectiveness of machine and deep learning for cyber security. DOI: 10.23919/CYCON.2018.8405026
- [18] Naway, A.; LI, Y. A Review on The Use of Deep Learning in Android Malware Detection.International Journal of Computer Science and Mobile Computing, Vol.7 Issue.12, December- 2018, pg. 42-58. Available online: http://arxiv.org/abs/1812.10360 (accessed on 3 April 2022).
- [19] UCI Machine Learning Repository, "Phishing Websites Dataset" https://archive.ics.uci.edu/ml/datasets/phishing+websites
- [20]Ozcan, A., Catal, C., Donmez, E. et al. A hybrid DNN–LSTM model for detecting phishing URLs. Neural Comput & Applic 35, 4957–4973 (2023). doi.org/10.1007/s00521-021-06401-z