24th International Conference on Information Technology, Computer and Telecommunication
Event Place: Tbilisi,Georgia
www.itcconf.ir
بیست و چهارمین کنفرانس بین المللی فناوری اطلاعات، کامپیوتر و مخابرات | گرجستان
24th International Conference on Information Technology,Computer and Telecommunication

Phishing Attacks Detection a Machine Learning-Based Approach

Niloofar Sasannia

School of Industrial and Information engineering Polytechnic university of Milan Telecommunications engineering, Milan, Italy

Abstract

Machine learning methods are more effective than simple blacklisting strategies in detecting phishing attacks due to their ability to adapt to new types of attacks and not requiring manual modification. However, the selection of features and classifiers for these methods directly affects the detection performance. Therefore, in this study, we carefully analyze the contribution of various features and products to the detection of phishing attacks in order to find the best as a different effectiveness measure, including the latest, newest, best F1 score, and best F1 score technique. Using a good strategy, each combination of certain groups is divided into various classifications to identify phishing websites. In addition to our data, two existing datasets were used for further analysis. Test results show that Uniform Resource Locator (URL) and Hypertext Transfer Protocol (HTTP) based features provide the best performance, if not all. It is one of the fastest distributions, outperforming other distributions with an F1 score of 0.99.

Keywords:Machine learning, Phishing, Classifications, Detection

1. Introduction

Currently, most companies and organizations use digital tools to provide fast and easy access to services we use in our daily lives. However, this also brings with it data security issues. Personal information, financial information and passwords used to access these services can create information security issues. There are many different types of cyber attacks used to obtain personal and financial information. Phishing attack is one of these attacks that involves creating a fake website that copies a legitimate website and is known to steal the user's personal and financial information[1]. Phishing websites are created by copying and creating similar versions of legitimate websites. Therefore, victims have no doubt about the website they are accessing, because there is very little difference between a phishing site and a legitimate website. Support IT Security is an organization that studies the effects of phishing attacks[2]. Although the attacker's motivations and attack strategies are known, effective methods to prevent phishing attacks have not yet been developed. In the third quarter of 2022, the APWG observed a total of 1,270,883 phishing attacks, setting a new record, making it the worst phishing quarter the APWG has ever observed[3]. Meanwhile, APWG member OpSec Security found that phishing attacks targeting financial institutions, including banks, remained the most common type of attack, accounting for about 23% of all phishing attacks. Attacks on webmail and software as a service remained stable, accounting for about 17% of all attacks, while attacks on retail/e-commerce sites fell to 4% from 14% in 2019. Phishing attacks targeting social media companies In the fourth quarter of 2021, 8 percent of all attacks were made, rising to 15 percent in the second quarter of 2022, before falling to 11 percent in the third quarter [2,3]. Phishing websites are also spread by posting on social media platforms or sending messages to victims on social networks[4]. Techniques to identify attacks. Most of these tactics are blacklisting[1], which involves creating a list of malicious websites known to be involved in phishing attacks. When a user tries to visit a website, the system can check the author of the website against the blacklist and block access if the website appears on the list. Blacklisting can be a great way to block known phishing websites, but it's not

foolproof. New phishing sites are being created all the time, and it's hard to keep a blacklist up to date. Attackers will also sometimes use names or websites that aren't on the blacklist, making their attacks harder to detect and stop. Machine learning involves the use of algorithms that can learn from data, make predictions, or make decisions [5,6]. Phishing log information and legitimate websites. The model will be able to learn characteristics commonly associated with phishing attacks and use that information to predict how likely a new website is to be a phishing attack. They're more useful than blacklists because they can adapt to new phishing attacks and don't need to update the blacklist. However, feature selection and classification algorithms have a direct impact on intrusion detection, so wrong selection of features and classifiers will degrade performance. Therefore, this study carefully analyzed the contribution of various features and classification algorithms towards the detection of phishing attacks in order to find the best classifiers and feature set that will improve the detection. For this purpose, we prepared a new dataset and made it available to the research community via https://github.com/sibelkapan/phishing_dataset (Access date: October 10, 2023). This file contains 500 phishing and 500 legitimate websites, and contains 25 Uniform Resource Locators (URLs), Hypertext Markup Language (HTML), and Hypertext Transfer Protocol (HTTP) attributes. Using a well-chosen strategy, each combination of these groups is fed into various classification systems including nearest neighbor (k-NN), support vector machine (SVM), Naive Bayes (NB), decision tree (DT), Multilayer Perceptron. (MLP) and Stochastic Gradient Descent (SGD) [5,6], Long phishing website. Five different performance metrics including accuracy, precision, recall, F1 score and time distribution were used while evaluating the performance of different features and classifiers [5,6]. During the experiment, two existing datasets were used in addition to new training data. Therefore, the main contributions of our work can be summarized as follows: finding the best classifier for phishing attack detection based on the effectiveness of the features and different metrics.

Data and methods In this section, we introduce our new data, explain its features, and briefly explain the classification and performance metrics. A part of this work has been published in Kapan's Master's thesis[39]. Data and features In our study, we prepare new data to train and evaluate machine learning models to detect phishing websites. The dataset is available to the research community at https://github.com/sibelkapan/phishing_dataset (accessed October 10, 2023). While preparing the data, URLs of legitimate and phishing websites were collected by Alexa and PhishTank platforms, respectively. , also manages many websites, including the most used platforms such as popular search engines and services, social media platforms, financial websites, shopping sites, etc. In addition, websites are further classified according to their content. Therefore, the rankings and analytics provided by the platform only direct users to legitimate websites. Since phishing attacks have a major impact on financial and commercial transactions, we use the terms "financial" and "property" to write URLs related to these topics. In addition, websites with the terms "login" and "update" were added to thedataset to include websites related to login and transaction functions, which are frequently used in phishing attacks. Therefore, the legal group of the dataset includes not only legitimate websites but also lower-ranked websites. Legitimate websites that can be confused with phishing sites are also included in the dataset.

2.Classifier

A classifier or classification algorithm is a machine learning algorithm designed to classify input data into predefined classes or classes [5,6]. The goal of the classification algorithm is to determine the relationship that suggests strategies for class objectives based on the training process. Activities are represented by related features. After training, the algorithm can predict new classes without seeing the situation. Therefore, in our study, the input data corresponds to websites and features such as URL, HTML, and HTTP, because the features are extracted from the input source. As mentioned earlier, the target group can

be phishing or legitimate. Six named classifiers were used to classify websites in our study: SVM, k-NN, DT, SGD, NB, and MLP [5,6]. The effectiveness of these products in combination with the above methods has been investigated and compared for each test.] This is done by finding the hyperplane with the largest edge, which is called the vector, which is the distance between the hyperplane and the closest data in each class. In this way, SVM can be extended well for invisible objects. In our test setup, we use the constant 1 and the radial basis function (RBF) kernel. It works by storing all the available information and when new information needs to be classified or valued, it looks at the k closest points (based on some distance metric) and returns the rank or value of the k nearest neighbors. k is user-defined, can be thought of as a parameter. One of the main advantages of K-NN is its simplicity because it requires very little training, unlike many other classification algorithms. In our experimental setup, we consider the optimal value of k to be 5,6] units. The idea here is to create a tree-like model of decisions and their outcomes. Thus, at each point in the tree, a decision is made based on the value of one of the inputs, and each page represents a list. Or Price, depending on the job. In our experimental setup, we use Gini impurity as the segmentation technique.

3.Performance metrics

We use five different performance metrics (such as accuracy, precision, recall, F1-score, and deployment time) to measure the effectiveness of the search discovery method. These metrics are important for understanding the model's strengths and weaknesses in distinguishing between good (phishing) and bad (legitimate) groups, as well as its speed at detecting challenges. Four values are included: true positive (TP), negative (FP), negative (TN), and false negative (FN). TP corresponds to cases where the model correctly predicts the class. So in the context of phishing site detection, a positive result occurs when a model is identified as a phishing site when it is actually a phishing site. FP corresponds to cases where the model does not correctly predict the class. In other words, the vulnerability occurs when a model mistakenly identifies a legitimate website as phishing. TN is the result of the model predicting the negative class. The real disadvantage is when the model identifies the website as legitimate and it is a legitimate website. FN is the result of the model failing to predict the negative class. In other words, false negatives occur when the model fails to identify a phishing website and correctly classifies it as a legitimate website. As shown in (1), it measures the proportion of correct predictions (true positives and negatives) for all events in the data. Accuracy represents the ability of the model to make correct predictions for both positive (phishing) and negative (legitimate) cases.

4.Results and Discussion

In the experimental study, the contribution of different devices and different methods to the detection performance of phishing websites was analyzed using the full search strategy. As mentioned before, files are characterized by URL, HTML and HTTP groups. An exhaustive search method [44] was used on these groups and the data was divided into seven subsets based on all possible group combinations: URL, URL + HTML, URL + HTTP, HTML, HTML + HTTP, HTTP and URL + HTML + HTTP runs. With the help of the links the effectiveness, relationships and uniqueness of different groups and classification algorithms are revealed. (TM) i5-2430M CPU @ 2.40 GHz and 8 GB RAM. Python programming language and Scikit-learn library [45] are used to implement all methods. The comparison data is shown in Table 2, where the best value of each feature is in bold. Therefore, the highest F1 score (0.99) was achieved by URL + HTTP feature set and DT classifier. On the other hand, URL or URL + HTTP features with NB classifier gives the lowest F1 score (0.53). The highest accuracy (0.99) was achieved using URL + HTTP feature set and DT classifier, while URL or URL + HTTP features using NB classifier gave the lowest accuracy (0.67). URL + HTTP feature set and DT classifier gave the highest accuracy (0.99), while

HTML feature set and NB classifier gave the lowest accuracy (0.68). When it comes to getting down to business, there are many leaders. URL + HTTP or HTTP or URL + HTML + HTTP feature set Using SVM classifier, HTTP feature set using SGD classifier, HTTP feature set Using NB classifier, URL + HTTP or HTTP feature set using MLP classifier > URL + HTTP feature set using K-NN classifier and HTTP feature set using DT classifier gave the best recovery (0.99). On the other hand, the lowest recovery rate (0.37) was achieved using URL or URL + HTML or URL + HTTP feature set of NB classifier. br> The fastest model has a split time of 0.01 seconds while HTML + HTTP feature set with MLP classifier is the slowest model with a split time of 1.00 seconds.

Conclusion

Phishing attacks pose a threat to the security of organizations and individuals. Machine learning can be a powerful tool to detect these attacks. Classifiers and features are an important part of this process. By selecting the most suitable classifier and evaluating the correlation between different features, the effectiveness of phishing detection can be increased. Explore the programs for phishing performance in terms of various metrics such as cost, accuracy, F1 score, and detection time. We have also released a new, publicly available phishing dataset for the research community. According to our evaluation, the URL + HTTP feature set with the DT classifier performed best, achieving the highest F1 (0.99), accuracy (0.99), and precision (0.99). The NB classifier with a larger feature set proved the lowest F1 score (0.53), accuracy (0.68), and precision (0.67). In addition, the recovery performance of different components varies; SVM consistently provides the best recovery (0.99) in most cases. Classification time analysis shows that DT classifier not only achieves the best F1 score but also one of the shortest classification times, while NB is the fastest in the URL set (0.01 s), while MLP is the slowest in the HTML + HTTP feature set (1.00 s). Additional tests on other benchmarks confirm the performance of DT and SVM, making them reliable options for phishing detection. It is important to build a robust and effective phishing system for phishing attacks. Based on the results of our research, we can conclude that organizations can improve their ability and speed to detect and respond to phishing attacks and protect themselves better by using machine learning and appropriate users from this Threat Feature Set. Cost analysis can also be performed considering the effects of fast recovery.

Refrences

[1] Asiri, S.; Xiao, Y.; Alzahrani, S.; Li, S.; Li, T. A survey of intelligent detection designs of HTML URL phishing attacks. IEEE Access 2023, 11, 6421–6443. [CrossRef]
[2] APWG Anti-Phishing Working Group. Available online: https://apwg.org (accessed on 10 October 2023).
[3] APWG Phishing Activity Trends Report Q3. 2022. Available online: https://apwg.org/trendsreports (accessed on 10 October 2023).
[4] Tinubu, C.O.; Falana, O.J.; Oluwumi, E.O.; Sodiya, A.S.; Rufai, S.A. PHISHGEM: A mobile game based learning for phishing awareness. J. Cyber Secur. Technol. 2023, 7, 134–153. [CrossRef]
[5] Bishop, C.M. Pattern Recognition and Machine Learning; Springer: Berlin/Heidelberg, Germany, 2006. Zhou, Z.H. Machine Learning; Springer Nature: Berlin/Heidelberg, Germany, 2021.
[6] Khonji, M.; Iraqi, Y.; Jones, A. Phishing detection: A literature survey. IEEE Commun. Surv. Tutor. 2013, 15, 2091–2121. [CrossRef]

[7] Mohammad, R.M.; Thabtah, F.; Mccluskey, L. Tutorial and critical analysis of phishing websites methods. Comput. Sci. Rev. 2015, 17, 1–24. [CrossRef]
[8] Google Safe Browsing API. Available online: https://developers.google.com/safe-browsing/v4 (accessed on 10 October 2023).
[9] Netcraft Anti-Phishing Toolbar. Available online: https://www.netcraft.com/apps (accessed on 10 October 2023).
[10] Whittaker, C.; Ryner, B.; Nazif, M. Large-scale Automatic Classification of Phishing Pages. In Proceedings of the 17th Network & Distributed System Security Symposium, San Diego, CA, USA, 28 February–3 March 2010; pp. 1–14.
[11] Jain, A.K.; Gupta, B.B. A survey of phishing attack techniques, defence mechanisms and open research challenges. Enterp. Inf. Syst. 2022, 16, 527–565. [CrossRef]
[12] Qabajeh, I.; Thabtah, F.; Chiclana, F. A recent review of conventional vs. automated cyber-security anti-phishing techniques. Comput. Sci. Rev. 2018, 29, 44–55. [CrossRef]
[13] Moore, T.; Clayton, R.; Stern, H. Temporal Correlations between Spam and Phishing Websites. In Proceedings of the 2nd USENIX Workshop on Large-Scale Exploits and Emergent Threats, Boston, MA, USA, 21 April 2009; pp. 1–8.
[14] Thomas, K.; Grier, C.; Ma, J.; Paxson, V.; Song, D. Design and Evaluation of a Real-Time URL Spam Filtering Service. In Proceedings of the IEEE Symposium on Security and Privacy, Oakland, CA, USA, 22–25 May 2011; pp. 447–462.
[15] Gangavarapu, T.; Jaidhar, C.D.; Chanduka, B. Applicability of machine learning in spam and phishing email filtering: Review and approaches. Artif. Intell. Rev. 2020, 53, 5019–5081. [CrossRef]
[16] Zhang, Y.; Hong, J.; Cranor, L. CANTINA: A Content Based Approach to Detecting Phishing Web Sites. In Proceedings of the 16th International Conference on World Wide Web, Banff, AB, Canada, 8–12 May 2007; pp. 639–648.
[17] Wardman, B.; Stallings, T.; Warner, G.; Skjellum, A. High-Performance Content Based Phishing Attack Detection. In Proceedings of the eCrime Researchers Summit, San Diego, CA, USA, 7–9 November 2011; pp. 1–9.