BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2023
http://excel.fit.vutbr.cz

Malicious Domain Detection from External Data Sources

Adam Horák

Abstract
This paper aims to analyze available sources of information about internet domains to help improve methods of detecting malicious domain names. We collect a wide range of data from various sources for a rich data set. We extract the most insightful features from this set and train a classifier. This work aims to improve the accuracy and reliability of domain name classification models. We show that combining data from external sources only can already accurately predict malicious domains.

xhorak69@vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Malicious actors use domains for many harmful purposes, such as phishing, malware distribution, and botnet attacks. Thus, developing accurate and reliable methods for detecting these domains is essential to protect users from online threats.

We present our analysis of various sources of information about internet domains to improve the accuracy and reliability of domain name classification models. We collected a rich dataset from external sources such as DNS, TLS, and RDAP. We then extracted the most insightful features from this dataset to train a classifier using XGBoost. Our current focus is mainly on detecting phishing domains, which significantly threaten online users [1].

Contribution We analyzed which external sources provide helpful information for domain malignancy classification. On experimental results, we show concrete features that best boost the accuracy.

2. Related Work

Detecting malicious domains has been extensively studied in recent years. Prior work falls into three distinct categories. Some researchers focus on lexical features to classify domains as benign or malicious [2, 3]. Others analyze webpage content to identify malign domains [4, 5, 6]. The third common way uses external data sources, such as WHOIS and TLS [7, 8], similar to our approach.

However, these approaches rarely combine more than a few available data types and features [9, 10]. In contrast, our study combines many data sources and feature engineering techniques to maximize the variety of information used to detect phishing domains effectively.

3. External Data Sources

External data plays a critical role in our approach to detecting malicious domains. DNS records provide crucial information, such as the domain's name servers, mail servers, and IP addresses. RDAP is a WHOIS replacement that provides information on domain registration, including the registrar, registrant, and contact information, which may relate to known malicious activity. TLS is another crucial source of data that provides certificate chains used to verify the identity of a server. We can detect discrepancies in the chain. Geolocation and reputation data for IP addresses are also valuable sources of information. We can detect if the domain's IP addresses reside in a high-risk area or are associated with known malicious activities.

4. Classifier Creation

In this section, we will discuss our data collection and feature extraction process and how we train and evaluate our classifier using this data.

4.1 Data Collection

In order to create a rich dataset for training, we designed a program that loads domain lists and resolves relevant data for them. The program is extensible, allowing us to add new data sources and update our collections easily.

To ensure the quality of our dataset, we carefully selected our benign and malicious domain lists. We use the Cisco Umbrella top domains list for benign domains, which contains the most popular and reputable internet domains. We use various MISP feeds for malicious domains, which provide up-to-date information on known online threats.

4.2 Feature Extraction & Engineering

To use the data for classifier training, we must select insightful numeric values from this data and encode other information using feature engineering techniques to create new, informative features. To accomplish this, we developed a custom program that runs a modular pipeline of transformations on the data from our database. We picked and crafted quantifiable features from each source, amounting to over 40. The resulting tables then provide inputs for our classifier.

4.3 Model Training & Evaluation

To classify domains and aid in feature selection, we used the XGBoost framework. It is advantageous because it can report on feature importance, allowing us to identify the most informative features for our classification.

Our model training and evaluation process involved several steps. First, we evaluated the extracted features and iterated on them based on their importance, as reported by XGBoost. Our primary goal was to find a good set of features that could accurately and reliably classify domains as benign or malicious.

To evaluate our classifier's performance, we used k-fold cross-validation. We aimed for a high F1 score because our datasets are unbalanced, as we have more benign domains than malicious ones in our set.

5. Experimental Results

Our phishing classifier achieved an F1 score of 90% using external data only, with 0.95 precision and 0.87 recall for the phishing class. The confusion matrix in Table 1 shows false predictions on a test sample of 88,606 domains.

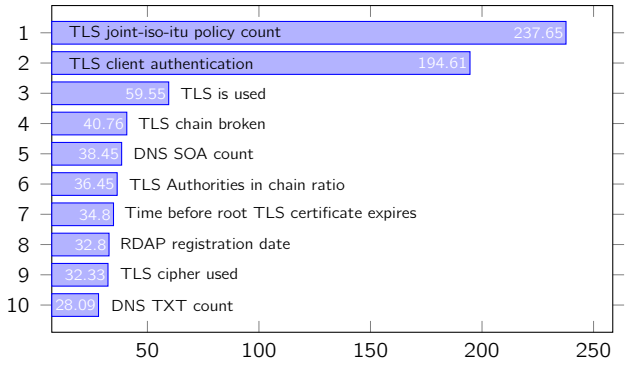

Figure 1. Features with highest gain values

		predicted	
		benign	phishing
true	benign	76950	558
	phishing	1450	9648

Table 1. Confusion matrix

The most important features were related to TLS and DNS, as shown in Figure 1. RDAP was often missing from phishing data and contributed little to our classifier's performance. This fact highlights the importance of using diverse data sources for domain analysis in various malign categories.

6. Conclusion

We demonstrated the effectiveness of combining external data sources for detecting malicious domains. Our approach achieved a high F1 score for phishing detection, primarily leveraging TLS-related features. These findings have important implications for developing more accurate and reliable methods of malicious domain detection.

External data sources can also combine well with lexical classifiers that rely on analyzing the textual content of domain names. Integrating external data with lexical analysis should further improve the accuracy of domain classification models. Future research can explore the combination of these different approaches.

Acknowledgements

I want to thank my supervisor Ing. Radek Hranický, Ph.D., for his help. The same goes for my fellow teammates. The research is supported by "Flow-based Encrypted Traffic Analysis" project, no. VJ02010024 granted by Ministry of the Interior of the Czech Republic and "Smart information technology for a resilient society" project, no. FIT-S-23-8209 granted by Brno University of Technology.

References

[1] Mohamed Alsharnouby, Furkan Alaca, and Sonia Chiasson. Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies*, 82:69–82, 2015.

[2] Khulood Al Messabi, Monther Aldwairi, Ayesha Al Yousif, Anoud Thoban, and Fatna Belqasmi. Malware detection using dns records and domain name features. ICFNDS '18, New York, NY, USA, 2018. Association for Computing Machinery.

[3] Hong Zhao, Zhaobin Chang, Weijie Wang, and Xiangyan Zeng. Malicious domain names detection algorithm based on lexical analysis and feature quantification. *IEEE Access*, 7:128990–128999, 2019.

[4] John McGahagan, Darshan Bhansali, Ciro Pinto-Coelho, and Michel Cukier. A comprehensive evaluation of webpage content features for detecting malicious websites. In *2019 9th Latin-American Symposium on Dependable Computing (LADC)*, pages 1–10, 2019.

[5] Mario Heiderich, Tilman Frosch, and Thorsten Holz. Iceshield: Detection and mitigation of malicious websites with a frozen dom. In Robin Sommer, Davide Balzarotti, and Gregor Maier, editors, *Recent Advances in Intrusion Detection*, pages 281–300, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[6] Dongjie Liu and Jong-Hyouk Lee. Cnn based malicious website detection by invalidating multiple web spams. *IEEE Access*, 8:97258–97266, 2020.

[7] Leyla Bilge, Engin Kirda, Christopher Kruegel, and Marco Balduzzi. Exposure: Finding malicious domains using passive dns analysis. In *Ndss*, pages 1–17, 2011.

[8] Yong Shi, Gong Chen, and Juntao Li. Malicious domain name detection based on extreme machine learning. *Neural Processing Letters*, 48(3):1347–1357, July 2017.

[9] Masahiro Kuyama, Yoshio Kakizaki, and Ryoichi Sasaki. Method for detecting a malicious domain by using whois and dns features. In *The third international conference on digital security and forensics (DigitalSec2016)*, volume 74, 2016.

[10] Amirreza Niakanlahiji, Bei-Tseng Chu, and Ehab Al-Shaer. Phishmon: A machine learning framework for detecting phishing webpages. In *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 220–225, 2018.