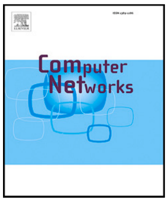
Computer Networks 255 (2024) 110847

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet



MC-Det: Multi-channel representation fusion for malicious domain name detection☆

Yabo Wang [a], Ruizhi Xiao [a], Jiakun Sun [a], Shuyuan Jin [a,b,c,*]

[a] *School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, 510000, China*
[b] *Guangdong Provincial Key Laboratory of Information Security Technology, Guangzhou, 510000, China*
[c] *MoE Key Laboratory of Information Technology, Sun Yat-sen University, Guangzhou, 510000, China*

ARTICLE INFO

Keywords:
Domain Name System
Malicious domain detection
Network topological structure
Multi-channel embedding representation
Resource associations

ABSTRACT

As the essential fundamental infrastructure of the current network, the Domain Name System is widely abused by cyber attackers, malicious domain detection has become a crucial task in combating cyber crime. Most existing methods focus on local attributes, treating each domain name individually. Alternatively, they prioritize global associations among domain names, but ignore the attributes of the domains themselves, allowing malicious domain names to survive through sophisticated evasion techniques. In this paper, we propose MC-Det, a hybrid framework for detecting malicious domain names by fusing a Multi-channel representation of domain names. MC-Det first abstracts the domain name resolution process into three spatially independent information channels: Attribute space, which contains the intrinsic information in the domain name string itself, Constraint space, which involves the potential constraints imposed on the network activity behind the domain name, Topological space, which represents the actual usage and deployment of the domain name. Subsequently, it generates proper embedding representations of domain names for each channel. This novel Multi-channel representation provides a comprehensive understanding of domain name resolution process. Finally, a Multi-channel fusion strategy employing by attention mechanism is used to generate the final representation of domain names for the classifier, making MC-Det suitable for malicious domain name detection in different application scenarios. Experimental results demonstrate that MC-Det outperforms other state-of-the-art techniques, while only utilizing the resource information revealed in the domain name resolution phase.

1. Introduction

The Domain Name System (DNS) is the central component of the current Internet, it maps human-friendly domain names to numeral IP addresses. In most cases, resolving domain names is the initial step when users connect to cyber resources. Consequently, DNS has been extensively abused by cyber attackers to carry out malicious campaigns, including phishing [1,2], botnets [3,4], spam [5,6], malware [7,8], etc. These malicious campaigns have resulted in significant financial losses. Hence, malicious domain name detection becomes the first line of defense in protecting user privacy and property.

To address these issues, a denylist has been proposed to block malicious domain names, and it is still one of the most commonly used techniques currently. However, evasion techniques are also continuously evolving. Techniques such as Domain-flux, Fast-flux [9], Domain Generation Algorithms (DGAs) [10], and even Generative Adversarial Network (GAN) [11] have been employed to render malicious domains more covert and diversified, making it difficult for denylist to match the update rate of malicious domains. In order to effectively identify meticulously crafted domain names, numerous detection techniques based on machine learning have been proposed. These methods focus on the local attributes of domain names and rely on feature engineering to extract features that carefully designed by experts, which have proven to be effective. Nonetheless, all of them treat domain names individually, with the limitation of labor-intensive and feature-expired, making them susceptible to evasion by sophisticated techniques. Therefore more researchers have started to investigate mining potential malicious domain names with association between malicious domains

☆ This paper is supported by National Natural Science Foundation of China, Award Number: 62472456; Basic and Applied Basic Research Foundation of GuangDong Province, Award Number: 2022B1515120072; and Guangdong Provincial Key Laboratory of Information Security Technology, Award Number: 2023B1212060026.

* Corresponding author.
E-mail address: jinshuyuan@mail.sysu.edu.cn (S. Jin).

https://doi.org/10.1016/j.comnet.2024.110847
Received 19 March 2024; Received in revised form 22 August 2024; Accepted 3 October 2024
Available online 5 October 2024
1389-1286/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.