Computer Networks 259 (2025) 111073

Contents lists available at ScienceDirect

Computer Networks

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

Efficient DNS over HTTPS servers discovery method: A voting-based stacked ensemble model with secure connection metadata

ZunDong Zhang, Yanan Cheng *, Haiyan Xu, Zhaoxin Zhang *, Chao Li

Faculty of Computing, Harbin Institute of Technology, Harbin, 150001, China

ARTICLE INFO

Keywords:
DNS
DNS over HTTPS
Ensemble algorithm
DNS measurement
Cybersecurity

ABSTRACT

In today's rapidly developing internet landscape, user privacy and data security have become paramount concerns for both the public and enterprises. Traditional DNS queries transmitted in plaintext are susceptible to various security threats. DNS over HTTPS (DoH) emerged to address these concerns by encapsulating DNS queries within the HTTPS protocol, thereby enhancing the security and privacy of DNS queries. However, due to the complexity of DoH's configuration and the inefficiency of existing discovery methods, significant challenges remain in discovering and utilizing DoH servers. Most current methods for discovering publicly accessible DoH servers are time-consuming and resource-demanding and do not comprehensively gather information. To address these issues, this study proposes an Efficient active DoH Discovery Method based on Secure Connection Metadata, abbreviated as EDDM-SCM, improving the DoH discovery process to encompass both IP addresses and domain names of the public DoH servers. Specifically, the method extracts key features from TLS and HTTPS connection information and employs a voting-based stacked ensemble model (VBSEM) to construct a DoH server filtering mechanism. This approach addresses the challenge of positive sample scarcity and effectively prevents model overfitting. Experimental results demonstrate that this method can identify over 95% of DoH servers while improving time efficiency by at least 70%, significantly reducing network resource consumption. Our findings revealed over 20,000 DoH servers, providing a novel and effective solution for actively discovering public DoH servers. This facilitates the widespread adoption and decentralization of DoH services.

1. Introduction

The Domain Name System (DNS), as a crucial component of internet infrastructure, is responsible for domain name resolution. However, traditional DNS queries are transmitted in plaintext, making them susceptible to security threats such as man-in-the-middle attacks and DNS cache poisoning [1,2]. To address these issues, DNS over HTTPS (DoH) [3] was developed, encapsulating DNS queries within the HTTPS protocol to enhance security and privacy. Compared to other DNS encryption protocols, DoH is currently the most widely adopted. For instance, according to research by APNIC [4], approximately 13.7% of global DNS traffic is transmitted via the DoH protocol, and this proportion continues to grow. Consequently, research into various aspects of DoH applications has become a significant focus in recent years.

Despite its increasing adoption, the use of DoH remains far below that of traditional DNS. One of the main reasons for this disparity is the difficulty in discovering DoH servers [5]. Currently, the primary method for finding DoH servers is through search engine searches, which leads to several issues:

Centralization of DNS Services. Most users rely on the top search results, resulting in severe centralization of DoH services. Research has confirmed this issue [6], showing that centralization brings various risks and undermines the original purpose of the DNS public service system.

Limited Visibility. Search results represent only a small portion of DoH servers, making it difficult for industry professionals to fully understand the deployment of DoH, thereby hindering the optimization of DoH deployment strategies.

Therefore, active discovering of DoH servers is crucial. Active discovering can overcome the limitations of search engines by comprehensively discovering DoH servers across the network, identifying and excluding potentially unsafe DoH servers, and providing industry professionals with a more complete picture of DoH deployment. However, active discovering of DoH servers presents the following two challenges:

* Corresponding authors.
E-mail addresses: 22b903090@stu.hit.edu.cn (Z. Zhang), chengyn@hit.edu.cn (Y. Cheng), xuhy@hit.edu.cn (H. Xu), zhangzhaoxin@hit.edu.cn (Z. Zhang), 20B903094@stu.hit.edu.cn (C. Li).

https://doi.org/10.1016/j.comnet.2025.111073
Received 8 October 2024; Received in revised form 16 January 2025; Accepted 18 January 2025
Available online 27 January 2025