DNS TUNNELING ISSUES

Jack Lin

Computer Networks - COSC 4378

Department of CS/EE

Section I - Abstract

The domain name system (DNS) has emerged as a fundamental component of the Internet since its inception, facilitating the establishment of websites, file transfers, and email communications. DNS functions by directing users to the appropriate computers, applications, and files through the mapping of Internet Protocol (IP) addresses to domain names. However, the presence of various security vulnerabilities within DNS has rendered it susceptible to exploitation by malicious actors, who can create covert channels known as DNS tunnels. These tunnels enable unauthorized access to sensitive information and facilitate a range of attacks while evading detection by intrusion detection systems (IDS). Furthermore, attackers have leveraged DNS tunnels for clandestine communications, including the exfiltration of confidential data and the establishment of command and control (C2) infrastructures. Over time, DNS has increasingly become a focal point for cybercriminals, offering avenues for network breaches and data exfiltration. Improper configuration of DNS can expose systems to numerous threats, including DNS-based malware, amplification attacks, false positives, and tunneling exploits. In response to these challenges, Google and Cloudflare have introduced DNS over TLS (DoT) and DNS over

HTTPS (DoH) as protective measures against DNS-related attacks. These protocols are designed to enhance privacy and security by encrypting DNS traffic between users and DNS resolver servers. While DNS over HTTPS significantly bolsters user privacy and security, it simultaneously complicates the ability of network administrators to identify malicious traffic. To address DNS-related issues, various statistical and machine learning methodologies have been developed to detect and mitigate risks associated with DNS tunneling activities. The efficacy of these algorithms has been validated through comprehensive evaluations across diverse datasets and scenarios, underscoring their potential to strengthen cybersecurity initiatives. This paper seeks to elucidate the threats posed by DNS tunneling and to formulate effective strategies for mitigating these attacks, thereby equipping organizations with the necessary tools to safeguard their digital assets.

Keywords - Domain Name System (DNS), Top-level domains (TLDs), Resource Records (RRs), Cache poisoning. Domain Name System Security Extensions (DNSSEC), Zone signing keys (ZSKs), key signing keys (KSKs), DNS tunneling, Domain generation algorithm (DGA), DNS over HTTPS (DoH), Machine learning (ML), Anomaly detection, Explainable artificial intelligence (XAI)

Section II - Introduction

The Domain Name System (DNS) is a cornerstone of the Internet infrastructure, serving as a hierarchical, distributed database that translates human-readable domain names into machine-readable IP addresses. This essential function enables seamless access to websites, email, and online services. Despite its foundational role, DNS was originally designed without robust security features, leaving it vulnerable to attacks such as cache poisoning and

man-in-the-middle attacks. To address these issues, DNS Security Extensions (DNSSEC) were introduced, employing cryptographic signatures to ensure the authenticity and integrity of DNS responses. By establishing a chain of trust from the root zone through the DNS hierarchy, DNSSEC enhances security but also introduces complexities like key management and compatibility challenges. Attackers have found ways to exploit DNS for malicious purposes, such as through DNS tunneling, a covert technique for data exfiltration and command-and-control communication. By encoding data in DNS queries and responses, often using attacker-controlled DNS servers, DNS tunneling enables cybercriminals to steal sensitive information or maintain control of compromised systems. This activity is challenging to detect due to its resemblance to legitimate DNS traffic. Furthermore, the advent of DNS over HTTPS (DoH) adds another layer of complexity, encrypting DNS traffic to enhance privacy but making it harder for traditional security systems to identify malicious activities. Advanced techniques, including machine learning models and explainable AI, have emerged to analyze DNS traffic for detecting such threats, underscoring the evolving challenges and innovations in securing DNS as the backbone of Internet communication.

Section III - Background

The Domain Name System (DNS) constitutes a fundamental component of the Internet's infrastructure. It serves as a distributed and extensible database that converts human-friendly domain names, like example.com, into machine-readable IP addresses, such as 192.0.2.1. The architecture of DNS is hierarchical, beginning with the root zone at the apex. Beneath the root zone are top-level domains (TLDs), followed by second-level domains, such as example.com, subdomains like www.example.com, and individual hostnames. DNS employs a variety of resource records (RRs), including A records for IPv4 addresses, MX records for mail servers,

and CNAME records for domain aliases. These records are maintained in authoritative DNS servers, which hold the accurate data for designated domains. DNS servers are typically categorized as either authoritative or caching (recursive). Authoritative servers retain the official DNS records for a domain, while caching servers temporarily store query responses to alleviate the burden on authoritative servers. The communication protocol within DNS relies on structured DNS messages, as outlined in RFC 1035. These messages are divided into sections: a query section that solicits information about a domain, a response section that provides the results of the query, and an authoritative section that directs to authoritative servers for additional resolution. Despite its critical role in Internet navigation, DNS was initially developed without robust security features, rendering it susceptible to various attacks, including cache poisoning, man-in-the-middle attacks, and identity spoofing. There is no inherent mechanism to authenticate the validity of DNS responses. To tackle these challenges, the Domain Name System Security Extensions (DNSSEC) were developed to enhance the integrity and authenticity of DNS information. DNSSEC employs cryptographic signatures to confirm that DNS responses are genuine and unaltered. Utilizing public-key cryptography, where zone signing keys (ZSKs) authenticate DNS records and key signing keys (KSKs) validate the ZSKs, DNSSEC creates a trust chain extending from the root zone through the DNS hierarchy. This mechanism enables DNS resolvers to authenticate data by matching the signature against the corresponding public key. Although DNSSEC significantly bolsters security by reducing vulnerabilities to prevalent threats like cache poisoning, it also brings forth new challenges, including key management, signature expiration, and compatibility with existing DNS systems. For DNSSEC to function effectively, both authoritative DNS servers and recursive resolvers must support cryptographic validation, depending on trust anchors—usually the public key of the root zone—to authenticate

DNSSEC signatures throughout the DNS hierarchy. The process of DNS resolution is recursive, involving queries to multiple DNS servers, beginning with the root server, followed by the top-level domain (TLD) server, and concluding with the authoritative server for the specific domain. This multi-tiered approach ensures that users are directed to legitimate sites, thereby enhancing the security and reliability of DNS as a fundamental component of Internet communications. By fortifying DNS against attacks and confirming the authenticity of data, DNSSEC fosters greater trust in the system, addressing the escalating security requirements of contemporary Internet protocols. As the DNS infrastructure continues to develop, it is crucial to ensure that both authoritative and recursive servers adhere to DNSSEC standards to preserve the security and integrity of global domain name resolution.

Section IV - State-of-Art

DNS tunneling represents a stealthy technique employed by cybercriminals to exfiltrate data or retain control over compromised systems by leveraging the domain name system (DNS). The DNS is an integral component of the Internet's infrastructure, responsible for converting domain names into IP addresses, thereby facilitating user access to websites, email, and various online services. Given that DNS traffic is fundamental to nearly all Internet activities, it is typically not obstructed by firewalls, rendering it a favorable conduit for malicious actors. In the context of DNS tunneling, attackers embed data within DNS queries and responses, frequently utilizing encoding methods such as Base32 or Base64. This encoded information is transmitted to a DNS server under the attacker's control, which subsequently decodes the data, enabling the theft of sensitive information or the maintenance of control over an infected device. The process initiates when malware on the compromised machine generates DNS queries that contain concealed data, such as pilfered files or credentials. These queries are directed towards fraudulent domains, often

generated through a domain generation algorithm (DGA) that changes frequently to evade detection. Upon reaching the attacker's DNS server, these queries are decrypted, allowing for the retrieval of the stolen data. This covert communication method is challenging to identify, as it mimics standard DNS traffic, which is typically overlooked by most cybersecurity systems. Beyond data exfiltration, DNS tunneling can also facilitate other nefarious activities, including the command and control of botnets, malware distribution, and the execution of additional attacks on vulnerable systems. The inherent difficulty in detecting DNS tunneling arises from the critical role of DNS traffic in network operations; blocking such traffic could disrupt legitimate Internet access. Nevertheless, various strategies can be employed to identify and mitigate DNS tunneling, including traffic analysis to uncover atypical patterns, such as unusually large DNS queries or recurrent access to dubious domains. Advanced methodologies, including machine learning and deep learning algorithms, are capable of scrutinizing DNS traffic to uncover anomalies and detect tunneling activities by recognizing patterns that diverge from typical behavior. Furthermore, DNS over HTTPS (DoH) has emerged as a significant technology aimed at mitigating privacy issues by encrypting DNS queries transmitted over HTTPS. Nevertheless, the implementation of DoH presents new obstacles in the analysis of DNS traffic, particularly in the identification of malicious actions such as DNS tunneling. The DoH protocol safeguards the content of DNS queries from unauthorized access, thereby complicating detection efforts by conventional cybersecurity frameworks. This encrypted DNS traffic frequently mimics standard HTTPS traffic, which complicates the identification of harmful DNS tunnels concealed within. While DoH enhances user privacy, it simultaneously enables attackers to circumvent traditional network security protocols, including firewalls and intrusion detection systems (IDS). The detection of malicious DoH traffic is further complicated by the encrypted nature of the

communication, which obstructs the examination of the actual DNS query. Recent studies have employed machine learning (ML) techniques to tackle this challenge. Models such as Random Forest (RF), Gradient Boosting (GB), and XGBoost have been utilized to analyze DNS traffic for anomaly detection, even when encrypted through DoH. By concentrating on traffic attributes such as packet size, timing, and flow patterns, these models can effectively classify DNS traffic as either benign or malicious, even when transmitted over encrypted channels like DoH. These approaches are becoming increasingly vital as the adoption of DNS over HTTPS expands and malicious actors exploit this encryption for nefarious activities, including data exfiltration and command-and-control communications. One notable initiative in this domain is the CIRA-CIC-DoBre-2020 dataset, which encompasses both benign and malicious DNS over HTTPS (DoH) traffic. This dataset serves as a foundation for training machine learning models aimed at detecting anomalous patterns within DoH queries. Although these models have demonstrated enhanced accuracy, significant challenges remain in achieving optimal performance, particularly in the context of sophisticated encryption techniques and the intricate behaviors exhibited by threat actors. The integration of explainable artificial intelligence (XAI) methodologies, such as SHAP (SHapley Additive exPlanations), holds potential for increasing the transparency of machine learning models. Conventional machine learning frameworks are frequently perceived as "black boxes," complicating the understanding of the rationale behind their outputs. XAI strategies elucidate the features or traffic patterns influencing the model's predictions, thereby fostering trust among security practitioners in the model's findings and enabling informed responses. This level of transparency is crucial for enhancing the dependability of automated systems tasked with identifying DNS tunneling and various DNS-related attacks. A critical insight regarding DNS tunneling is the observed decline in cache

hit rates on DNS cache servers during such attacks. Typically, high cache hit rates are characteristic of normal network activities; however, during a DNS tunneling incident, the elevated volume of queries utilized for data exfiltration can lead to increased cache errors.

Section V - Conclusion

In summary, the Domain Name System (DNS) is a crucial component of Internet functionality, facilitating smooth interactions between users and diverse online services. By converting user-friendly domain names into machine-readable IP addresses, DNS simplifies the process of accessing websites and applications. Nevertheless, the significance of DNS is overshadowed by notable security weaknesses that present substantial risks to both individuals and organizations. A particularly alarming vulnerability is the potential for DNS tunneling, a method through which cybercriminals can embed data within DNS queries and responses. This technique can be leveraged for various malicious purposes, including data exfiltration, which involves the covert extraction of sensitive information from networks, and secret communications that allow attackers to establish hidden channels for command-and-control activities. The introduction of encryption technologies, such as DNS over HTTPS (DoH), further complicates the security environment by enabling these illicit actions to evade conventional security protocols, thereby making it more challenging for organizations to identify and address threats. In light of these issues, numerous innovations have been developed to bolster the security and privacy of DNS. Solutions like DNS Security Extensions (DNSSEC), DNS over TLS (DoT), and DoH have been introduced to offer additional safeguards against various cyber threats. While these advancements enhance the overall security framework of DNS, they also present new challenges for network administrators and cybersecurity strategies. The deployment of these technologies necessitates meticulous planning and oversight to prevent the inadvertent creation of new

vulnerabilities or disruption of legitimate traffic. To effectively address the diverse threats confronting DNS, organizations must implement a holistic security approach that incorporates multiple layers of defense. This approach must encompass the implementation of comprehensive security measures designed to safeguard DNS traffic against interception and manipulation. Moreover, the integration of machine learning and explainable artificial intelligence (AI) is pivotal for detecting anomalies, allowing organizations to recognize atypical behavioral patterns that could signify a security breach. Ongoing surveillance of DNS traffic is equally critical, as it enables the immediate identification of malicious activities and supports prompt action against potential threats. By utilizing these advanced techniques, organizations can significantly improve their capacity to detect and respond to harmful actions while also proactively mitigating the risks linked to DNS vulnerabilities. In addition, continuous research and development in DNS security are essential to adapt to the ever-changing landscape of cyber threats. As adversaries persist in devising new methods and tactics, it is crucial for organizations to maintain a vigilant and flexible stance regarding DNS security.

REFERENCES

[1] F. Zou, S. Zhang, B. Pei, L. Pan, L. Li and J. Li, "Survey on Domain Name System Security," *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)*, Changsha, China, 2016, pp. 602-607, doi: 10.1109/DSC.2016.96.

[2] IETF. RFC 1035 Domain Names - Implementation and Specification[DB/OL]. 1987-11[2011-8-16] http://www.ietf.org/rfc/rfc1035.txt.

[3] IETF. RFC 4033 DNS Security Introduction and Requirements[DB/OL]. 2005-3[2012-3-21] http://www.ietf.org/rfc/rfc4033.txt.

[4] G. Gürsoy, A. Varol and A. Nasab, "DNS Tunnel Problem In Cybersecurity," *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*, San Antonio, TX, USA, 2024, pp. 1-6, doi: 10.1109/ISDFS60797.2024.10527301.

[5] T. Zebin, S. Rezvy and Y. Luo, "An Explainable AI-Based Intrusion Detection System for DNS Over HTTPS (DoH) Attacks," in *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2339-2349, 2022, doi: 10.1109/TIFS.2022.3183390.

[6] S. K. Singh and P. K. Roy, "Detecting Malicious DNS over HTTPS Traffic Using Machine Learning," *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*, Sakheer, Bahrain, 2020, pp. 1-6, doi: 10.1109/3ICT51146.2020.9312004.

[7] N. Ishikura, D. Kondo, V. Vassiliades, I. Iordanov and H. Tode, "DNS Tunneling Detection by Cache-Property-Aware Features," in *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 1203-1217, June 2021, doi: 10.1109/TNSM.2021.3078428.