POSTER

An Overview of Threats Exploring the Confusion Between Top-Level Domains and File Type Extensions

ANDERSON SALES, Polytechnic University of Viana do Castelo, Viana do Castelo, Viana do Castelo, Portugal

NUNO TORRES, Polytechnic University of Viana do Castelo, Viana do Castelo, Viana do Castelo, Portugal

PEDRO PINTO, Polytechnic University of Viana do Castelo, Viana do Castelo, Viana do Castelo, Portugal

Published: 19 June 2024

Citation in BibTeX format

CODASPY '24: Fourteenth ACM Conference on Data and Application Security and Privacy
June 19 - 21, 2024
Porto, Portugal

Conference Sponsors:
SIGSAC

An Overview of Threats Exploring the Confusion Between Top-Level Domains and File Type Extensions

Anderson Sales
Instituto Politécnico de Viana do Castelo
Viana do Castelo, Portugal
ansales@ipvc.pt

Nuno Torres
Instituto Politécnico de Viana do Castelo
Viana do Castelo, Portugal
nunotorres@ipvc.pt

Pedro Pinto
Instituto Politécnico de Viana do Castelo and INESC TEC
Viana do Castelo and Porto, Portugal
pedropinto@estg.ipvc.pt

ABSTRACT

Cyberattacks exploit deceptions involving the Domain Name Service (DNS) to direct users to fake websites, such as typosquatting attacks, which exploit natural typographical errors, and homograph attacks, where different Unicode characters resemble the legitimate ones.

The deception attacks may also exploit the confusion between DNS domain names, specifically Top-Level Domains (TLDs), and file extensions. Recently, two new TLDs were added, "zip" and "mov", sharing names with certain file types. This overlapping can be explored by malicious actors in a range of threat scenarios to compromise user security.

This paper provides an overview of threats originating from the confusion between specific TLDs and file extensions, such as the recent "zip" and "mov". The threats are grouped into 6 threat scenarios that are described and discussed. This research can be part of a more comprehensive strategy that includes addressing the risks associated with these threats and designing future strategies to address the threats associated with exploiting this ambiguity.

CCS CONCEPTS

• **Security and privacy** → **Human and societal aspects of security and privacy**; **Web application security**; *Domain-specific security and privacy architectures*; • **Information systems** → *World Wide Web*.

KEYWORDS

Top-Level Domains, File Extensions, Security Threats, Cyber Security, Cyber Attack

ACM Reference Format:
Anderson Sales, Nuno Torres, and Pedro Pinto. 2024. An Overview of Threats Exploring the Confusion Between Top-Level Domains and File Type Extensions. In *Proceedings of the Fourteenth ACM Conference on Data and Application Security and Privacy (CODASPY '24), June 19–21, 2024, Porto, Portugal.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3626232.3658641

CODASPY '24, June 19–21, 2024, Porto, Portugal

ACM ISBN 979-8-4007-0421-5/24/06
https://doi.org/10.1145/3626232.3658641

1 INTRODUCTION

The Domain Name Service (DNS) is an important internet service that supports the translation of domain names into IP addresses. As with other services, the DNS is the target of multiple attacks aiming at its availability or stability.

Specific attacks exploit the deceptions involving the DNS service to direct users to fake websites, such as typosquatting attacks, exploring natural typographical errors commonly made when manually entering a Uniform Resource Locator (URL), or homograph attacks, which deceive users using different Unicode characters. An example of the latter is the publication of CVE-2018-4277 [8] that affected the Safari browser on Apple products, where a Latin-specific unicode (U+A771) was used to trick the end-user by successfully spoofing the letter "d" causing ambiguity in rendering [16].

A DNS deception attack can also exploit the confusion between Top-Level Domains (TLDs) and file extensions. TLDs are located at the top of the domain name hierarchy of DNS and provide a logical categorization of web addresses, identifying websites, and providing information about the nature and origin of online resources. Internet Assigned Numbers Authority (IANA) [4] defined that a subset of the TLDs, the Generic Top-Level Domains (gTLD), can take multiple forms and, in 2012, Internet Corporation for Assigned Names and Numbers (ICANN) opened an application window for the first application round accepting proposals of new TLDs [5].

More recently in 2023 [17], to expand options for users and businesses, Google proposed a set of new TLDs including ".zip" and ".mov" [17]. These new TLDs overlapping their name with known file extensions, can be the trigger of security threats [13].

This paper provides an overview of threats originating from the confusion between specific TLDs and file extensions, such as the recent "zip" and "mov".

The remaining document is organized as follows: Section 2 presents the related work. Section 3 details the threat scenarios. Section 4 presents the conclusions and future work.

2 RELATED WORK

The attacks exploiting deceptions in the DNS service, such as Internationalized Domain Name (IDN) homograph attacks, or typosquatting attacks, have been an active threat [10, 12].

A set of research works are focused on TLDs operators. In [3], the authors addressed DNSs security and stability through a control plane for TLDs operators. Also, this work provides insights into mitigating vulnerabilities in specific TLDs, including cases where identical file extensions may represent security risks.

The authors in [14] analyzed DNSs configurations and their security by examining resource records of TLDs. It is also discussed

that an analysis of DNSs configurations can provide valuable information on how to avoid ambiguities and associated security risks.

The article in [1] presented a tool to detect malicious domains using statistical IDN features in TLDs. While this work focuses on detecting malicious domains, it can offer insights into identifying potential TLDs with identical file extensions that attackers may exploit.

The confusion between file extensions and TLDs is also discussed in a set of research works. In [11], the authors explored the evolution and adoption of TLDs and Domain Name System Security Extensions (DNSSEC). In this research work the authors discussed how the evolution of TLDs can help to contextualize the issue of identical file extensions and their implications for cybersecurity.

In [2], authors highlighted the emergence of using file extensions as TLDs, presenting a possible future exploitation vector where such ambiguity can result in security loopholes.

In [6], the author concluded that new TLDs represent a serious cybersecurity risk and could be used to bypass security measures. The authors also recommended that organizations take security measures to protect themselves from these risks.

Additionally, the author in [7] underscored the significant concern the new scenario brings to cyberspace, further reinforcing the need to adopt respective countermeasures.

To our knowledge, no previous work enumerates the threats originated by the confusion between specific TLDs and file extensions, such as the recent "zip" and "mov".

3 THREAT SCENARIOS

The confusion between TLDs and file extensions may be explored in multiple attack forms. The threats and the set of attacks associated with this confusion reassemble in Open Redirects attacks, Phishing attacks, Supply Chain attacks, Cryptocurrency Mining Scripts, Domain Spoofing, and Social Media Spambots.

These attacks can be depicted as follows:

(1) Open Redirect attacks - an Open Redirect attack can be performed using a trusted TLDs to a malicious URL with ".zip" extension and deceive users. For instance, the legitimate website:
https://go.dev/linux-amd64.zip
may have an open redirect pointing to a malicious website by using the following:
https://go.dev@linux-amd64.zip
When checking the Uniform Resource Identifier (URI) parsing any element before the "@" will be ignored by the browser and will forward the end user to the malicious domain:
http://linux-amd64.zip

(2) Phishing - Phishing attacks are well-known for exploiting users' trust and leading them to malicious websites through deceptive URLs. In this scenario, phishing scams make use of the ".zip" TLDs to host websites that resemble legitimate download portals. Unsuspecting users can be tricked into believing that they are downloading an authentic ".zip" file when they are being redirected to a malicious website. Using the ".zip" TLDs helps mask malicious links as legitimate file downloads, thereby increasing the effectiveness of phishing attacks and the success rate of deceiving victims.

(3) Supply Chain Attack - Threat actors can compromise the supply chain of legitimate software, or service, by inserting malicious ".zip" files with misleading TLDs. When downloaded and executed by users, these files can introduce malware into systems as they are perceived as legitimate files. These attacks can have a high impact as happened in the SolarWinds case in [15], where the consequences of such attacks highly impacted organizations and their customers. After this severe cyber incident, it is estimated that the monetary loss was approximately $90M, as reported by BitSight [9].

(4) Cryptocurrency Mining Scripts - Hackers can host websites with the ".zip" TLDs that automatically runs cryptocurrency mining scripts on visitors' machines. These scripts can be disguised as part of downloading or unpacking files. The rationale for downloading a ".zip" file may serve as an explanation for the high processing consumption, typically associated with cryptocurrency mining, thereby reducing the likelihood that users will suspect and identify malicious activity.

(5) Domain Spoofing - In this attack, the actors create websites using the ".zip" TLDs, which are designed to closely resemble the legitimate domains of trusted organizations. The aim is to trick users into providing confidential information or performing malicious actions. Using the ".zip" TLDs on these spoofed domains can bring an additional appearance of legitimacy, making it more likely to trick unsuspecting users into trusting and interacting with the malicious website. For instance, a legitimate domain hosting a download file such as:
https://go.dev/dl/go1.20.1.linux-amd64.zip
can be spoofed to the following malicious website:
https://go.dev.dl.go1.20.1.linux-amd64.zip

(6) Social Media Spambots - Criminals can use websites with the ".zip" TLDs to host malicious content or links, which are then spread through social media spambots. These links may pose as "unique" downloads or "leaked" files to trick users into clicking on them. Using the ".zip" TLDs can make these links appear as harmless file downloads, increasing click-through rates, and the potential spread of malware.

The above-enumerated attacks explore a range of threat scenarios derived from the confusion between TLDs and identical file extensions. Identifying and categorizing these threats helps foster a proactive approach to enhance the security of systems and users. Understanding these attack vectors is fundamental to developing effective protection strategies and protecting digital assets regarding evolving cyber threats.

4 CONCLUSION

A set of deception attacks exploits the confusion between DNS domain names, such as typosquatting and homograph attacks. A subset of these attacks exploits the confusion between TLDs and file extensions. The introduction of two new TLDs, "zip" and "mov", enables exploiting these names' ambiguity by malicious actors, and

a range of threat scenarios to compromise user security can be enumerated.

This paper provided an overview of threats stemming from the confusion between specific TLDs and file extensions. The threats are grouped into 6 attacks, namely Open Redirects attacks, Phishing attacks, Supply Chain attacks, Cryptocurrency Mining Scripts, Domain Spoofing, and Social Media Spambots.

This research can be part of a more comprehensive strategy that includes addressing the risks associated with these threats and designing future strategies to address the threats associated with exploiting this ambiguity. Ultimately, the research in this area can strengthen user protection and enhance the reliability of their online interactions.

ACKNOWLEDGMENTS

This research was funded by Applied Digital Transformation Laboratory (ADiT-Lab), a Research and Development (R&D) unit, based at the Instituto Politécnico de Viana do Castelo (IPVC). This research was developed in the context of the Master in Cybersecurity of IPVC.

REFERENCES

[1] Alshaima Almarzooqi, Jawahir Mahmoud, Bayena Alzaabi, Arsiema Ghebremichael, and Monther Aldwairi. 2022. Detecting Malicious Domains Using Statistical Internationalized Domain Name Features in Top Level Domains. arXiv:2211.08020 [cs.CR]

[2] Joshua Aquino and Stephen Hilt. 2023. *Future exploitation vector file extensions as top level domains.* Trend Micro. Retrieved February 25, 2024 from https://www.trendmicro.com/en_us/research/23/e/future-exploitation-vector-file-extensions-as-top-level-domains.html

[3] Cristian Hesselman, Giovane C.M. Moura, Ricardo De Oliveira Schmidt, and Cees Toet. 2017. Increasing DNS Security and Stability through a Control Plane for Top-Level Domain Operators. *IEEE Communications Magazine* 55, 1 (2017), 197–203. https://doi.org/10.1109/MCOM.2017.1600521CM

[4] IANA. 2023. *IANA Root Zone Database.* IANA. Retrieved March 12, 2024 from https://www.iana.org/domains/root/db

[5] ICANN. 2012. *New Generic Top-Level Domains.* ICANN. Retrieved March 12, 2024 from https://newgtlds.icann.org/en/about/program

[6] Scott Ikeda. 2023. *New google top-level domains that use common file extensions could pose a serious cyber risk.* CPO Magazine. Retrieved March 01, 2024 from https://www.cpomagazine.com/cyber-security/new-google-top-level-domains-that-use-common-file-extensions-could-pose-a-serious-cyber-risk/

[7] Stan Kaminsky. 2023. *Security risks of the .zip and .mov domains.* Kaspersky Blog. Retrieved February 05, 2024 from https://www.kaspersky.com/blog/zip-mov-domain-extension-confusion/48254/

[8] National Vulnerability Database NIST. 2024. *NVD - CVE-2018-4277 — nvd.nist.gov.* NIST. Retrieved March 26, 2024 from https://nvd.nist.gov/vuln/detail/CVE-2018-4277

[9] Samit Shah. 2021. *The financial impact of Solarwinds breach.* Bitsight. Retrieved February 09, 2024 from https://www.bitsight.com/blog/the-financial-impact-of-solarwinds-a-cyber-catastrophe-but-insurance-disaster-avoided

[10] Ax Sharma. 2024. *Registrars can now block all domains that resemble brand names — bleepingcomputer.com.* Bleeping Computer. Retrieved March 26, 2024 from https://www.bleepingcomputer.com/news/technology/registrars-can-now-block-all-domains-that-resemble-brand-names/

[11] Yo-Der Song, Aniket Mahanti, and Soorya Charan Ravichandran. 2019. Understanding Evolution and Adoption of Top Level Domains and DNSSEC. In *IEEE International Symposium on Measurements & Networking (M&N).* IEEE, 1–6. https://doi.org/10.1109/IWMN.2019.8805011

[12] David Strom. 2024. *Typosquatting Wave Shows No Signs of Abating — darkreading.com.* Dark Reading. Retrieved March 26, 2024 from https://www.darkreading.com/threat-intelligence/typosquatting-wave-shows-no-signs-of-abating

[13] Johannes Ullrich. 2023. *The .zip gTLD: Risks and Opportunities.* SANS Internet Storm Center. Retrieved March 12, 2024 from https://isc.sans.edu/diary/The+zip+gTLD+Risks+and+Opportunities/29838

[14] Mengyuan Wang, Zhaoxin Zhang, and Haiyan Xu. 2017. DNS configurations and its security analyzing via resource records of the top-level domains. In *11th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID).* IEEE, 21–25. https://doi.org/10.1109/ICASID.2017.8285736

[15] Jake Williams. 2020. *What You Need To Know About the SolarWinds Supply-Chain Attack.* SANS Institute. Retrieved March 10, 2024 from https://www.sans.org/blog/what-you-need-to-know-about-the-solarwinds-supply-chain-attack/

[16] Tencent Security Xuanwu Lab XLAB. 2018. *Spoof All Domains Containing 'd' in Apple Products [CVE-2018-4277] — xlab.tencent.com.* Tencent. Retrieved March 26, 2024 from https://xlab.tencent.com/en/2018/11/13/cve-2018-4277/

[17] Christina Yeh. 2023. *Google Registry: 8 new top-level domains for dads, grads and techies.* Google Registry. Retrieved March 12, 2024 from https://blog.google/products/registry/8-new-top-level-domains-for-dads-grads-tech/