Threats on Top-Level Domains with Identical File Extensions

Anderson Sales[1,3], Nuno Torres[1][0000−0001−7598−3687], and Pedro Pinto[1,2][0000−0003−1856−6101]

[1] Instituto Politécnico de Viana do Castelo, 4900-347 Viana do Castelo, Portugal
[2] INESC TEC, 4200-465 Porto and Universidade da Maia, 4475-690 Maia, Portugal
[3] Corresponding Author: ansales@ipvc.pt

Abstract. With the increasing number of Top-level Domains (TLDs) being introduced, the potential for security risks and vulnerabilities also increases. A particular concern arises when ambiguity is created between TLDs and file extensions, leading to potential confusion and security threats. This article aims to explore the dangers posed by TLDs with identical file extensions and propose mitigation strategies.
Additionally, it was analyzed the potential vulnerabilities of TLDs with identical file extensions, illustrating how attackers can exploit them to compromise user security.
This paper underscores the significance of understanding and addressing the risks associated with TLDs that share identical file extensions.

Keywords: TLDss · file extensions · security risks · threat detection · cyber security · cyber attack.

1 Introduction

TLDs are fundamental elements of the Internet's domain name structure. They are located at the top of the domain name hierarchy and provide a logical categorization of web addresses. Each TLDs has a specific set of rules and restrictions defined by the organization responsible for administering it. Furthermore, they play a crucial role in identifying and differentiating websites by providing information about the nature and origin of online resources.

How TLDs work is described in Request for Comments (RFCs) 1951 [4], 1034 [6] and 1035 [7]:

RFC 1951: "Domain Name System Structure and Delegation": This RFC sets out general guidelines for the structure and delegation of domain names, including TLDs.

RFC 1034 and RFC 1035: "Domain Names - Concepts and Facilities" and "Domain Names - Implementation and Specification", respectively: These RFCs define the fundamental concepts related to domain names, including the hierarchical structure of TLDs.

In turn, file extensions are suffixes added to file names to indicate their type or format. They provide important information about how a file should be

interpreted and processed by the operating system or application. For example, the extensions ".docx" and ".pdf" indicate Microsoft Word document files and Portable Document Format (PDF) files, respectively. By correctly identifying file extensions, systems can apply appropriate actions, such as opening the file with the appropriate application or taking security measures.

The first appearance of ambiguity between file extensions and TLDs occurred when Microsoft launched DOS in 1981 which, in turn, had files with ".com" extensions within its system, and years later the .com TLDs was created in 1985; Over the years, the time has shown that such a coincidence was not a disaster from a cybersecurity perspective.

Nowadays, cyber security has become a critical concern as cyber threats are constantly evolving. Targeted attacks, phishing, malware distribution, and exploiting vulnerabilities are just a few examples of threats that can compromise the integrity, confidentiality, and availability of systems and data.

To mitigate these threats, it is necessary to understand and be aware of the risks associated with TLDs that use the same file extensions.

This article is organized as follows. Section 2 presents the related work. Section 3 explores different threat scenarios related to web security and file manipulation. Section 4 addresses the conclusions and future research.

2 Related Work

Although there are no previous studies directly related to this topic, it was found relevant research in adjacent areas, such as the work of the authors in [2] when discussing the security of TLDs and also the article of the authors in [5] which provides a detailed overview of how suspicious Uniform Resource Locator (URL) identification works. These studies provide an important context for understanding the general issue and help us to inform this work.

Among the research conducted, the works of the authors in [1] stand out, which address the detection of malicious domains in top-level domains. Also noteworthy is the authors' article in [3], which provides an overview of Domain Name Services (DNSs) attacks involving TLDs.

The next chapter consists of an analysis of threat scenarios and the presentation of specific solutions to address the identified challenges by describing the threats and vulnerabilities associated with the overlap between TLDs and file extensions, as well as possible solutions and mitigation strategies.

3 Vulnerability and Threat Scenarios

This section explores a set of vulnerabilities and threat scenarios arising from the ambiguity between TLDs and file extensions. The overlap between these two elements can create security loopholes and expose systems and users to significant risks. Understanding these scenarios is crucial for developing effective protection strategies.

1. Open Redirect attacks - it is possible to combine the Open Redirect attack with trusted TLDs to a malicious URL with .zip extensions and deceive users. For instance, the legitimate website hxxps[:]//go[.]dev/dl/linux-amd64[.]zip may have an open redirect pointing to a malicious website like: hxxps[:]//go[.]dev/dl/@linux-amd64[.]zip ; when checking the Uniform Resource Identifier (URI) parsing was noticed that any element before the "@" will be ignored by the browser and will forward the end user to the malicious domain linux-amd64[.]zip .
2. Phishing - 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, in reality, 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. Malware Delivery - In the context of TLDs and file extensions, malware delivery involves the use of deceptive elements to direct users to malicious websites, where malware is delivered and implanted on users' systems. In this scenario, attackers exploit the ambiguity between TLDs and file extensions to create fake URLs that redirect users to fraudulent websites. By visiting these sites, users may be tricked into downloading or executing malicious files, resulting in their systems being infected.
4. 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.
5. Cryptocurrency Mining Scripts - Hackers can host websites with the .zip TLDs that automatically run cryptocurrency mining scripts on visitors' machines. These scripts can be disguised as part of the process of downloading or unpacking files. The rationale for downloading a .zip file may serve as an explanation for the high CPU consumption typically associated with cryptocurrency mining, thereby reducing the likelihood that users will suspect and identify malicious activity.
6. Domain Spoofing - In this kind of attack threat actors can 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 add an additional appearance of legitimacy, making it more likely to trick unsuspecting users into trusting and interacting with the malicious website. For example, a legitimate domain hosting a download file like hxxps[:]//go[.]dev/dl/go1.20.1.linux-amd64[.]zip can easily be spoofed to the following malicious website: hxxps[:]//go[.]dev.dl.go1.20.1.linux-amd64[.]zip
7. Shortcut Disguise Attack - This kind of attack involves creating shortcuts or links that appear to be legitimate .zip files. These shortcuts are designed to trick users into believing they are opening a .zip file, but actually direct

them to a malicious website when clicked. This site may contain various threats such as malware downloads or phishing pages. The strategy behind this attack is to exploit users' familiarity with .zip files and the confidence they have in opening these files. By disguising the shortcut as a .zip file, attackers increase the likelihood that users will fall into the trap, thereby compromising their security.

8. Data Exfiltration - Threat actors can utilize .zip TLDs as part of their command and control infrastructure, allowing data exfiltration from compromised systems by making it appear as if that data is being uploaded to a harmless .zip file. Using the .zip TLDs assists in concealing the whereabouts of stolen data, making it difficult for cybersecurity defenses to detect and block. This tactic aims to bypass protective measures and allow attackers to discreetly obtain confidential information.
9. Drive-By Downloads - Cybercriminals can take advantage of the .zip TLDs to perform automated downloads, in which a website downloads a malicious file onto a user's system without their knowledge or consent. Using the .zip TLDs can make these malicious downloads appear more legitimate, making them more difficult for users and security software to detect and block. This technique seeks to exploit users' trust when accessing websites that appear to be harmless, resulting in the silent infiltration of malware.
10. 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 in order to trick users into clicking on them. Using the .zip TLDs can make these links appear to be harmless file downloads, which increases click-through rates and the potential spread of malware.

4 Conclusion

In this study, we explored threat scenarios related to web browsing security and file manipulation in order to identify potential risks and vulnerabilities.

It is important to acknowledge that restricting TLDs as part of a security strategy has gained prominence in the current landscape. However, it is necessary to consider the practical implications and limitations associated with this measure.

In conclusion, the constantly evolving threat landscape demands that we adapt and adopt innovative approaches to protect users and their sensitive information during web browsing.

Some potential areas for future research include evaluating security measures and conducting a sampling analysis of malware and phishing attacks that utilize TLDs with file extensions to gain a better understanding of the specific threats associated with this aspect.

By addressing these areas, future research can contribute to expanding knowledge and improving security practices on the web, while addressing specific challenges. These investigations have the potential to strengthen user protection and enhance the reliability of their online interactions.

References

1. Almarzooqi, A., Mahmoud, J., Alzaabi, B., Ghebremichael, A., Aldwairi, M.: Detecting malicious domains using statistical internationalized domain name features in top level domains (2022)
2. Hesselman, C., Moura, G.C., De Oliveira Schmidt, R., Toet, C.: Increasing dns security and stability through a control plane for top-level domain operators. IEEE Communications Magazine **55**(1), 197–203 (2017). https://doi.org/10.1109/MCOM.2017.1600521CM
3. Householder, A., Houle, K., Dougherty, C.: Computer attack trends challenge internet security. Computer **35**(4), sulp5–sulp7 (2002). https://doi.org/10.1109/MC.2002.1012422
4. Jon Postel: Domain Name System Structure and Delegation. RFC 1591, Internet Engineering Task Force (1994), https://tools.ietf.org/html/rfc1591
5. Ma, J., Saul, L.K., Savage, S., Voelker, G.M.: Identifying suspicious urls: an application of large-scale online learning. In: Proceedings of the 26th annual international conference on machine learning. pp. 681–688 (2009)
6. P. Mockapetris: Domain Names - Concepts and Facilities. RFC 1034, Internet Engineering Task Force (1987), https://www.rfc-editor.org/rfc/rfc1034
7. P. Mockapetris: Domain Names - Implementation and Specification. RFC 1035, Internet Engineering Task Force (1987), https://www.rfc-editor.org/rfc/rfc1035