Poster: BEYONDGLYPH: Systematically Exploring and Exploiting Overlooked Attack Surfaces in Domain Name Homoglyph

Fasheng Miao[¶], Xiang Li[†✉], Hongyu Lin[¶], Changqing An[¶✉], Xiuqi Lu[¶],
Feng Zhang[¶], Chao Liu[¶], and Jilong Wang[¶‡]

[¶]*Institute for Network Sciences and Cyberspace, Tsinghua University*,
[†]*College of Cryptology and Cyber Science, Nankai University*, [‡]*Quan Cheng Laboratory*

***Abstract*—Domain names anchor user trust online. Homoglyph attacks exploit gaps between machine-level string storage and human visual perception to impersonate legitimate services. Yet, existing mitigations remain inherently *character-centric*, stagnating on simple single-character substitutions.**

We introduce BEYONDGLYPH, a framework bypassing modern defenses by exploiting overlooked domain representation properties: structural semantics, multi-character composition, invisible rendering control, bidirectional ordering, and encoding manipulation. Through real-world registration experiments and cross-platform evaluations, we instantiate 5 novel attack variants capable of visually spoofing arbitrary domains under major TLDs such as *.com*, *.net*, and *.org*. Crucially, two architectural blind spots drastically amplify this threat: the total lack of label validation in subdomain services, and the absence of URL sanitization in AI ecosystems. We demonstrate that web-enabled Large Language Models (LLMs) blindly ingest deceptive URLs, acting as unwitting proxies. Our large-scale evaluation across 26 browsers, 12 social media apps, and 17 LLMs reveals a catastrophic, ecosystem-wide security gap, underscoring the urgent need for domain name governance.

1. Introduction

Domain names serve as the primary anchor of user trust online, making them prime targets for homoglyph attacks that exploit the semantic gap between machine-level string storage and human visual perception [1]. Historically, research and defenses have focused overwhelmingly on single similar-character substitutions. Because modern mitigations are inherently *character-centric*, they remain blind to broader structural exploits. Consequently, there is a critical lack of systematic adversarial analysis regarding attack vectors beyond simple character substitution under modern Unicode and domain name standards.

Our Study. To bridge this gap, we systematically deconstruct the *end-to-end* domain resolution and rendering pipeline. Through RFC standard analysis, real-world registration experiments, and cross-platform evaluation, we introduce BEYONDGLYPH, a novel framework that systematically bypasses modern defense pipelines. Specifically, we instantiate 5 previously unexplored attack vectors: *(1)* structural semantic punctuation, *(2)* pure-ASCII multi-character visual fusion, *(3)* invisible rendering control injection, *(4)* bidirectional ordering, and *(5)* encoding prefix manipulation.

Crucially, through a systematic analysis of the registration logic across 64 domain services and the web-retrieval mechanisms of 17 mainstream LLMs, we uncover two architectural blind spots that drastically amplify this threat. First, subdomain provisioning services (e.g., CloudDNS) completely lack label-level security checks, serving as unregulated launchpads that bypass registry-level validations. Second, within AI ecosystems, modern Large Language Models (LLMs) equipped with web-retrieval capabilities blindly ingest deceptive URLs. By bypassing traditional browser UI warnings, these models act as unwitting proxies that "launder" phishing payloads to end users.

Contributions. In summary, our core contributions are threefold: *(i)* we conceptualize and instantiate 5 novel homoglyph variants capable of visually spoofing arbitrary domains; *(ii)* we expose catastrophic validation bypasses in subdomain ecosystems and LLMs, demonstrating real-world exploitability across 26 browsers, 12 social applications, and 17 LLM-based systems; and *(iii)* we responsibly disclose these findings to affected vendors, proposing actionable mitigations for this ecosystem-wide vulnerability.

2. The BEYONDGLYPH Methodology

Unlike legacy visual spoofing that relies on naive morphological similarity, which is now largely mitigated by mixed-script detection and Punycode warnings, BEYONDGLYPH systematically exploits the fundamental parsing, rendering, and encoding standards of the domain ecosystem.

Threat Model. As illustrated in Figure 1, we instantiate 5 unexplored attack vectors. Under this model, we assume an attacker only requires the capability to register domains and access target services. With these capabilities, attackers can deceive users and automated systems across three distinct boundaries: *(1) Browsers:* Evading standard UI safeguards. *(2) Apps:* Exploiting the absence of address bars for frictionless redirection. *(3) LLMs:* Abusing headless web-retrieval pipelines as unwitting proxies to launder malicious payloads without UI validation. The five novel attack variants are detailed below.

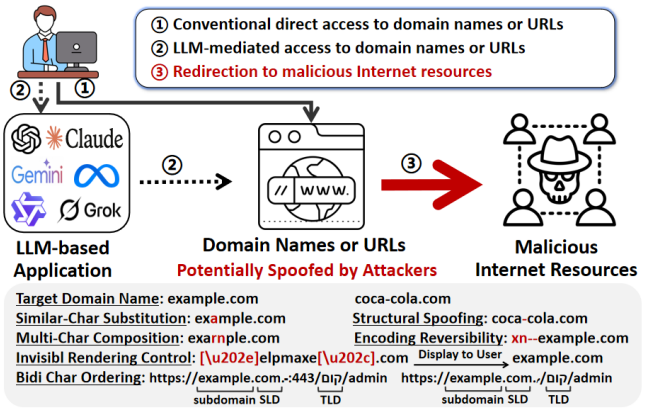


Figure 1: Threat Model of the **BEYONDGLYPH** Attack.

(1) V_{SChar} (Structural Semantic): Exploits delimiters (e.g., U+2024 ONE DOT LEADER) to make the SLD of *www.google.com* parse as *www.google*, bypassing alphanumeric filters.
(2) $V_{CharCom}$ (Multi-Char Composition): Substitutes characters with valid ASCII sequences (e.g., "*(r+n)rn*" for "*m*") via visual fusion, circumventing all script-mixing rules.
(3) $V_{ConChar}$ (Invisible Control): Injects hidden Unicode controls (e.g., U+202E RTL Override) to alter rendering (e.g., *elgoog.com* visually renders exactly as *google.com*).
(4) $V_{CharOrd}$ (Bidirectional Ordering): Abuses RTL TLDs. For example, registering *com.<RTL-TLD>* flips the visual order, allowing attackers to spoof arbitrary *.com* targets.
(5) $V_{PunyAlg}$ (Encoding Prefix): Registers standard ASCII domains prefixed with *xn–* (e.g., *xn–google.com*) to weaponize legitimate Punycode identifiers.
Discussion. Unlike legacy visual spoofing that relies on simple morphological similarity, BEYONDGLYPH introduces a novel framework by systematically exploiting foundational parsing, rendering, and encoding standards. This crucial shift from easily detectable character substitutions to system-level structural exploitation allows it to entirely bypass modern mixed-script defenses.

3. Real-World Impact & Exploitation Insights

We evaluated BEYONDGLYPH across 26 browsers, 12 social apps, and 17 LLMs (Windows, macOS, iOS, Android with the latest versions). Systems are vulnerable if they resolve malicious domains or retrieve content without explicit security warnings (e.g., Punycode fallback).
The Direct-Access Scenario (Browsers & Apps). Client-side defenses exhibit systemic failures. *All 26 browsers* are vulnerable to at least one variant; $V_{CharCom}$ and $V_{CharOrd}$ deceive 26 and 21 browsers, respectively. Alarmingly, several mobile browsers hide the address bar, rendering classic Punycode defenses obsolete. Additionally, 9 of 12 social apps seamlessly route users to malicious domains without warnings, enabling frictionless credential theft.
The LLM-Mediated Scenario. LLMs represent a massive, unchecked attack surface. *All 13 evaluated models with web retrieval* (e.g., ChatGPT) are vulnerable. Unlike traditional browsers, LLM retrieval pipelines rely on headless fetchers that lack visual UI safeguards. Consequently, they blindly resolve BEYONDGLYPH URLs, implicitly trusting the fetched content and incorporating malicious external context into trusted responses to effectively launder phishing payloads. Crucially, because this vulnerability stems from underlying structural parsing and infrastructure flaws, mitigations must be implemented at the system level (e.g., via strict URL validation proxies in the retrieval pipeline) rather than attempting to patch the issue through LLM model retraining.
Root Cause Analysis & Exploitation Gaps. Inspecting the Chromium rendering engine and 68 registration services (34 SLD + 34 subdomain providers) revealed two critical exploitation gaps:
(1) Registration Validation Bypass: While blocking mixed-script homoglyphs, registrars fail to restrict pure-ASCII compositions ($V_{CharCom}$) and invisible controls ($V_{ConChar}$). Furthermore, 29 of 34 subdomain services (e.g., CloudDNS) perform *zero* label-level validation. These unregulated launchpads allow attackers to bypass TLD registry safeguards and scale attacks effortlessly.
(2) Client-Side Blind Spots: Modern browser defenses overfit on single-character skeleton similarity and mixed-script detection. They remain fundamentally blind to structural semantics, bidirectional rendering logic, and encoding prefix manipulation, allowing BEYONDGLYPH domains to render identically to legitimate targets.
Responsible Disclosure: A total of 17 vendors have acknowledged our findings. Among them, OpenAI confirms that our findings highlight an insightful and highly relevant security concern for LLM-mediated browsing.

4. Conclusion

Moving beyond traditional character substitutions, we introduce the BEYONDGLYPH framework to systematically exploit domain structural semantics, multi-character composition, invisible controls, bidirectional ordering, and encoding prefixes. Our evaluation reveals a severe, ecosystem-wide security gap, exacerbated by unregulated subdomain services and web-enabled LLMs. Following responsible disclosure, we recommend moving from isolated character filters to holistic, system-level URL validation and UI sanitization across both traditional Web and GenAI ecosystems.
Acknowledgment. Authors from Nankai University were supported by the National Natural Science Foundation of China (No. 62502236, No. U25B2025), the Natural Science Foundation of Tianjin (No. 24JCQNJC02070), and the Open Project of National Engineering Laboratory for Technology of Internet Domain Name (No. KF202516). Authors from Tsinghua University were supported by the National Key Research and Development Program of China under Grant No.2020YFE0200500.

References

[1] E. Gabrilovich and A. Gontmakher, "The homograph attack," *Communications of the ACM*, 2002.

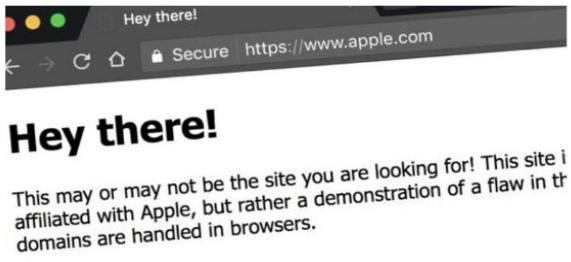
BEYONDGLYPH: Systematically Exploring and Exploiting Overlooked Attack Surfaces in Domain Name Homoglyph

Fasheng Miao [1], Xiang Li [2]✉, Hongyu Lin [1], Changqing An [1]✉, Xiuqi Lu [1], Feng Zhang [1], Chao Liu [1], ang Jilong Wang [13]
[1] Tsinghua University [2] Nankai University [3] Quan Cheng Laboratory

BEYONDGLYPH: Beyond Traditional Homoglyphs

- **Paradigm Shift**: Modern defenses are purely character-centric, missing broader structural and rendering exploits
- **BEYONDGLYPH**: Introduces 5 novel attack vectors that systematically bypassing modern visual & script-mixing defenses
- **Architectural Blind Spots**: Free subdomains acting as unregulated launchpads, web-enabled LLMs blindly serving as proxies to "launder" phishing payloads
- **Catastrophic Real-World Impact:** Arbitrary domain spoofing across 55 real-world systems, including 13 evaluated LLMs (e.g., ChatGPT)

Background



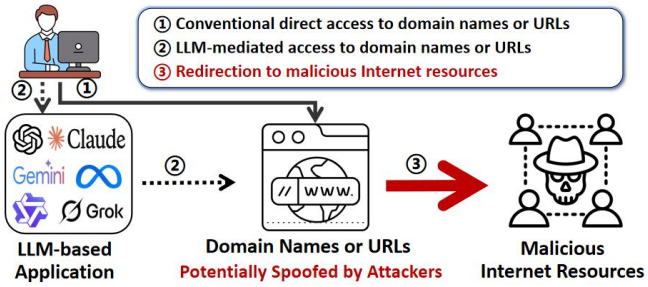
Traditional homoglyphs: Historically effective, but fundamentally limited

- Over-rely on single-char substitutions
- Prohibited by mod-ern registry policies
- Easily trigger modern IDN warnings
- Highly Detectable by Confusable datasets
- Ignore structural and rendering mechanics

Attack Model

Direct-Access: Evading browser UIs & mobile visual defenses.

LLM-Mediated:blindly fetch URLs, acting as unwitting proxies



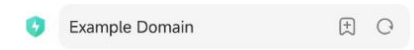
Attack Variants: Exploiting Structural Semantics & Rendering Mechanics

Target Domain Name: example.com — coca-cola.com
Similar-Char Substitution: example.com — Structural Spoofing: coca-cola.com
Multi-Char Composition: exarnple.com — Encoding Reversibility: xn--example.com
Invisibl Rendering Control: [\u202e]elpmaxe[\u202c].com —Display to User→ example.com
Bidi Char Ordering: https://example.com.-:443/קום/admin — https://example.com./קום/admin

subdomain SLD TLD subdomain SLD TLD

Attack Impact:

- **29/34 subdomains lack validation**
- **26/26 browsers & 9/12 apps compromised**
- **100% of web LLMs "launder" phishing URLs**
- **100% of Top-1M domains vulnerable**



Concealing the domain or URL

Hidden address bars amplify the threat

mfs24@mails.tsinghua.edu.cn