1859 Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Electrical, Electronics and Communications Engineering
(36thcycle)

Generative Approaches to Sound-Squatting: AI Tools and Validation

By

Rodolfo Vieira Valentim

******

Supervisor(s):
Prof. Marco Mellia, Supervisor
Prof. Idilio Drago, Co-Supervisor

Doctoral Examination Committee:
Prof. Tanja Zseby, Referee, Technische Universität Wien
Prof. Sebastian Garcia, Referee, Czech Technical University in Prague

Politecnico di Torino
2024

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Rodolfo Vieira Valentim
2024

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Generative Approaches to Sound-Squatting: AI Tools and Validation

Rodolfo Vieira Valentim

Cyber-squatting, a cyber-crime involving the registration of domain names to exploit established trademarks or identities, encompasses various strategies, including sound-squatting. Sound-squatting leverages phonetic similarities to deceive users, a risk amplified by the proliferation of smart speakers, voice assistants, and audio content. Sound-squatting presents challenges due to the inherent variations in pronunciation across different languages and among individuals. This thesis explores the complexities of sound-squatting across single-language, and cross-language scenarios.

Our primary goal is to develop a robust methodology for generating sound-squatting candidates that effectively handles various linguistic scenarios. By employing a data-driven approach using machine learning models, particularly Transformer Neural Networks, we successfully produce sound-squatting candidates across diverse scenarios, surpassing traditional list-based models in capturing pronunciation nuances.

The research follows a multi-stage process, starting with a naive baseline model and advancing to sophisticated architectures utilizing raw audio, spectrograms, and token-based pronunciation encoding. Evaluations are both quantitative and qualitative, assessing homophone coverage, quasi-homophone generation quality, and multi-language support.

Furthermore, the thesis examines whether these models can predict user transcription errors resulting from pronunciation misunderstandings. We compare collected user data on transcription mistakes with the model outputs to evaluate predictive accuracy.

Practical implications are explored through case studies on domain registrations and Python package repositories. We analyze Transport Layer Security (TLS) certificates and the Python Package Index (PyPI) to identify sound-squatting candidates, revealing real-world exploitation potential.

Key contributions of this thesis include developing a comprehensive approach to sound-squatting generation using advanced machine learning techniques, providing a systematic validation framework for assessing these tools, and conducting an analysis of the tools' predictive capabilities regarding user transcription errors. Additionally, the thesis offers an extensive checking of the sound-squatting attack surface in domain registrations and Python packages.

This research enhances the understanding of sound-squatting and provides practical tools and methodologies to mitigate associated risks. The findings highlight the need for proactive cybersecurity measures and offer valuable insights for future studies and applications.