NED UNIVERSITY JOURNAL OF RESEARCH

IDENTIFICATION OF PHISHING URLS BASED ON DISTINCTIVE LEXICAL AND TYPOSQUATTING PARAMETERS

Author(s): Kostyantyn Malyshenko[1] and Nikita Igorevich Potapovich[2]

Abstract:
Phishing attacks have become increasingly prevalent, with fraudsters seeking to obtain sensitive user data. This study introduces a novel approach for identifying critical parameters in URL classification, combining lexical features and indicators of typosquatting attacks. The methodology was tested on three datasets, including one designed to simulate typosquatting for research. The focus on both lexical and typosquatting parameters enhanced model speed but slightly reduced quality, with an average Area Under the Curve (AUC) reduction of 0.025 across datasets. Importantly, this approach does not depend on third-party evaluations, allowing for a comprehensive set of lexical occurrences. Experimental results were promising, achieving an F1 score of 93.2% with optimized parameters compared to 93.5% unoptimized, and 90.4% for the unoptimized balanced dataset versus 89.1% for optimized. The study concludes that incorporating both lexical parameters and typosquatting indicators is crucial for effective detection.

Keywords: Typosquatting, URL, Machine learning, Detection, Parameters

For full paper, contact:
Prof Muhammad Imran Aslam
Editor-in-Chief, NED University Journal of Research
Ph: +92 (21) 99261261-8 Ext:2670; Fax: +92 (21) 99261255
Email: NED-Journal@neduet.edu.pk
Website: http://www.neduet.edu.pk/NED-Journal

[1] Associate Professor, Department of Economics and Finance, V.I. Vernadsky Crimean Federal University, Republic of Crimea, Simferopol, Russia, Ph. +79788036860, Email: malyshenkoka@cfuv.ru
[2] Master's Student, Department of Economics and Finance, V.I. Vernadsky Crimean Federal University, Republic of Crimea, Simferopol, Russia, Ph. +79789818120, Email: n.potapovich@mail.ru