FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Assignment of bachelor's thesis

Title:	Malicious URL Detection in Real Network Traffic Using Machine Learning Methods
Student:	Vladimír Vávra
Supervisor:	Ing. Jaroslav Hlaváč
Study program:	Informatics
Branch / specialization:	Artificial Intelligence 2021
Department:	Department of Applied Mathematics
Validity:	until the end of summer semester 2025/2026

Instructions

Develop a model for malicious URL detection that achieves a sufficient balance between inference speed and quality, enabling deployment in real-world environments.

Theoretical Part:
1. Study and summarize the topic of malicious URL detection.
2. Conduct a literature review of standard techniques (e.g., regular expressions) and deep learning-based methods (e.g., convolutional, transformer-based approaches) used to address this problem.
3. Describe methods for model compression aimed at improving the speed/quality inference ratio.

Practical Part:
1. The student will compare all mentioned methods on the same datasets in terms of standard ML metrics (e.g., precision, F1-score, etc.), inference speed given the same computational resources, and speed/quality inference ratio. Comparisons will be based on the student's experiments and results published in related research papers.
2. Using the findings from previous research, the student will train a model with a sufficiently good speed/quality inference ratio to enable deployment in real-world environments. The evaluation will be done both on public datasets from [1].

Literature:

Electronically approved by Ing. Magda Friedjungová, Ph.D. on 18 January 2025 in Prague.

FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

[1] Liu, R., Wang, Y., Guo, Z., Xu, H., Qin, Z., Ma, W. and Zhang, F., 2024. TransURL: Improving malicious URL detection with multi-layer Transformer encoding and multi-scale pyramid features. Computer Networks, 253, p.110707.
[2] Maneriker, P., Stokes, J.W., Lazo, E.G., Carutasu, D., Tajaddodianfar, F. and Gururajan, A., 2021, November. Urltran: Improving phishing url detection using transformers. In MILCOM 2021-2021 IEEE Military Communications Conference (MILCOM) (pp. 197-204). IEEE.

Electronically approved by Ing. Magda Friedjungová, Ph.D. on 18 January 2025 in Prague.