PhD Opportunities in Hardware and AI Security

Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava (FIIT STU)

We invite highly motivated candidates to apply for PhD study at FIIT STU in the broad area of **hardware and AI security**. The available topics span neural network security, AI-assisted physical cryptanalysis, and the security of large language models.

Available PhD Topics

1. Hardware Security of Neural Networks

Neural networks are increasingly deployed on dedicated hardware such as microcontrollers, FPGAs, and ASIC accelerators to meet strict latency, energy, and cost requirements. While this enables efficient edge intelligence, it also introduces new hardware-level security risks. Physical implementations may unintentionally leak sensitive information or may be actively perturbed by an attacker.

This PhD topic focuses on the confidentiality and integrity of neural network implementations under **side-channel attacks** and **fault injection attacks**. The goal is to study novel attack strategies and develop practical countermeasures that remain compatible with the constraints of real-world deployments.

2. AI-assisted Physical Attacks on Cryptographic Implementations

Physical attacks on cryptographic implementations remain one of the most powerful practical threats to embedded security. Two major classes are **side-channel attacks**, which exploit physical leakage such as power, timing, or electromagnetic emanations, and **fault injection attacks**, which deliberately induce computation errors to reveal secret information.

This PhD topic investigates how artificial intelligence and machine learning can improve physical cryptanalysis. The project will study the current state of the art, propose new AI/ML-based methods for side-channel and fault analysis, and experimentally evaluate them on modern cryptographic targets such as AES, PRESENT, and selected post-quantum schemes.

3. Security of Large Language Models

Large language models (LLMs) are now being used in enterprise automation, software engineering, decision support, and human–computer interaction. As their adoption grows, their security becomes increasingly important. Two particularly relevant threat classes are **jailbreak and prompt-based attacks** and **physical/hardware-adjacent attacks** affecting the model-serving infrastructure.

This PhD topic focuses on the security of modern LLM systems, with emphasis on adversarial prompting, policy bypass, information leakage, integrity failures, and hardware-adjacent threats. The aim is to identify new failure modes, develop realistic attack models, and build a reproducible framework for evaluating the security of representative LLM deployments.

Who Should Apply?

We welcome applicants with strong motivation and a solid background in one or more of the following areas:

- computer security or applied cryptography,

- machine learning or trustworthy AI,
- embedded systems, digital design, or computer architecture,
- mathematics, data analysis, or algorithm design.

Programming experience in Python and/or C/C++ is highly desirable. Prior exposure to hardware platforms, deep learning frameworks, or security evaluation is a plus, but not strictly required.

How to Express Interest

Interested candidates are encouraged to contact us with:

- a short statement of research interests,
- a CV,
- and an indication of the preferred PhD topic.

Contact: Assoc. Prof. Xiaolu Hou
Email: houxiaolu.email@gmail.com
Web: xiaoluhou.github.io

The exact scope of each topic can be adjusted depending on the candidate's background, skills, and research interests.