Bachelor and Master Thesis Topics: Privacy, Security, and Digital Forensics

Chair of Parallelism & VLSI

February 11, 2026

1 Introduction

This document presents thesis topics in the areas of privacy-preserving technologies, anonymity networks, and digital forensics. These topics address current challenges and offer students opportunities to contribute to cutting-edge research in these fields. This document is part of the open thesis topic for the summer semester 2025 and structured in multiple parts. The first Section 2 details the thesis process and application requirements. Please read these carefully. Especially, the latter is helpful for a successful thesis topic application.

2 Thesis Procedure and Timeline

2.1 Application and Topic Assignment Process

Application and Topic Allocation All submitted applications will be reviewed and evaluated. Successful applicants will be offered specific thesis topics based on their qualifications, interests, and supervisor capacity. Students then need to confirm their topic selection.

Introduction Meeting All students who have been assigned topics from this document will attend a mandatory introduction meeting. This session will cover the general expectations and guidelines for thesis work.

Abstract Following topic assignment, students enter the abstract development phase with a maximum duration of 3 months for master students and 2 months for bachelor students. During this phase, students must:

1. **Literature Review:** Conduct research into related work and current state of the field
2. **Gap Analysis:** Identify relevant research gaps in accordance with the chosen thesis topic
3. **Research Questions:** Formulate specific research questions that address identified gaps
4. **Methodology Development:** Design appropriate methods and approaches to answer the research questions

Throughout the abstract development and following thesis working phase, students can regularly ask questions and discuss with their supervisor. Especially during the initial phase, questions and discussions are encouraged to refine the research approach.

Abstract Review Upon submission, the developed abstract will undergo a quality check to ensure feasible, well-defined research questions, an appropriate methodology and a realistic scope for the thesis timeline. If improvements are necessary, students will have the opportunity to refine and resubmit their abstract.

Official Registration Once the abstract is approved, the thesis will be officially registered with the university. The final submission deadline will be set according to university regulations and students receive confirmation of their registration and timeline.

A dedicated Moodle forum will serve as the central communication platform and provides additional resources and materials assisting you.

Students have flexibility in language choice:

- **Thesis Language:** English or German
- **Supervision Language:** English or German
- **Kolloquium Language:** English or German (independent of thesis language)

2.2 Application Requirements

To ensure the best match between students and thesis topics, please provide the following information in your application. This will help us assess your preparedness and align your interests with the most suitable research direction.

2.2.1 Required Application Components

Relevant Experience Describe relevant work experience, internships, research projects, or personal projects that could help with your chosen thesis topic. If you have no directly relevant experience, please explain any transferable skills or related activities.

Academic Background List university courses you have completed that provide helpful knowledge for your chosen topic.

Research Interest Statement In 2-5 sentences, explain what aspects of your chosen thesis topic you find most interesting and why. Consider addressing what specific challenges or problems within the topic motivate you and what unique perspective or approach you might bring to the topic.

Research Proposal Outline Provide a short text (approximately 200–400 words) that outlines

- From the provided list or based on your own interests, identify 2-3 specific research questions you would like to explore for your selected topic. Explain why these questions are important and how they contribute to the field.
- Describe your initial thoughts on how you would approach investigating these research questions. This should include general methods you plan to use for investigating the topic. How you plan to validate your findings or measure success? Do you have preliminary ideas for overcoming potential challenges?

Note: This is not a binding commitment but demonstrates your understanding of the research process.

Publication and Open Access Research results from these thesis topics may be of general interest to the privacy and digital forensics communities. We may seek to publish outstanding work through academic papers, technical reports, or open-source releases to make findings publicly accessible. This could be done in a form of openly published theses, open-source tools or academic papers. For academic papers, students remain the first author unless the work is combined with more comprehensive results from the supervisor. Please indicate if you are interested in a potential publication of the research results.

2.2.2 Application Format

Please structure your application as follows:

1. **Header:** Include your name, student ID, degree program (Bachelor/Master), and preferred thesis topic(s)

2. **Relevant Experience:** (maximum 300 words)
3. **Academic Background:** (Maximum 100 words)
4. **Research Interest Statement:** (2-5 sentences)
5. **Research Proposal Outline:** (200-400 words)
6. **Publication and Open Access:** (1-3 sentences)

Please submit your application via email to *joerg.keller@fernuni-hagen.de* (and CC to *pascal.tippe@fernuni-hagen.de*) with a **single PDF file attachment** containing all the components listed above. Alternatively, you may format the components as a Markdown file and send it attached as a plain text file

3 LLM Overblocking

The proliferation of powerful Large Language Models (LLMs), often accessed via cloud-based APIs, presents significant opportunities across various domains. However, it also affects our information access as the LLMs can be used for short inquiries, assisting with work tasks or gathering information. With few centralized providers, the content moderation becomes critical. While current research mostly looks at *hate speech* and *toxic content* in regards to LLL safety, another interesting venue is whether legal content is censored as well as this would be an imposed moderation by the model operator rather than a collectively agreed standard. As of today, there is no systematic study into which legal content is overblocked and how the model responds to these blocked inquiries.

LLM-Block: The goal of this topic is to assess whether and to what extent legal content is blocked by common LLMs.For this topic, students must acquire ground truth data and develop a framework to generate potentially sensitive prompts. These will be ingested into LLMs to categorize the responses.

4 Advancing Privacy in Cloud-Based LLM Interactions

The proliferation of powerful Large Language Models (LLMs), often accessed via cloud-based APIs, presents significant opportunities across various domains. However, this paradigm introduces substantial privacy risks, particularly when user prompts contain personally identifiable information (PII). The transmission of sensitive data to third-party servers raises concerns about potential misuse, unauthorized access, and compliance with increasingly stringent data protection regulations. A promising approach based on recent work from Grötzner and Tippe to mitigate these risks involves a hybrid or dual-LLM architecture, wherein a locally deployed LLM acts as a privacy-preserving intermediary. The fundamental concept underpinning this research is a system architecture featuring a local LLM, potentially lightweight and optimized for on-device execution, functioning as a *privacy guardian.* This local LLM intercepts user prompts intended for a more powerful cloud-based LLM. Its primary responsibilities include identifying PII within the prompt, applying a sanitization technique to remove or transform this PII, and then forwarding the modified, privacy-enhanced prompt to the cloud LLM for primary processing. Upon receiving a response from the cloud LLM (which is based on the sanitized prompt), the local LLM undertakes the crucial task of restoring the original PII into the response in a coherent and contextually appropriate manner before it is presented to the user. This approach seeks to harness the advanced capabilities of cloud LLMs while significantly reducing the exposure of raw PII to external entities.

LLM-Privacy The goal of this topic is to extend currently unpublished work from Grötzner and Tippe. This could include:

- How can sanitized answers be restored locally?

- How can local models be fine tuned to run a locally hosted privacy guardian?
- How can the privacy guardian handle multiturn-conversations or attachments?

5 Misusage of LLMs: Agentic Cybercrime

The transition from chatbots to agents like OpenClaw, Nanobot, and Goose marks a paradigm shift in the cyber threat landscape. While traditional LLMs are restricted by cloud-based safety filters, local agents operate with direct shell access, persistent memory, and the ability to execute code autonomously. This carries the risks that cybercrime actors increasing utilize them to improve their workflows or that LLMs even lower the barrier to enter into cybercrime.

LLM-Cybercrime: The goal of this topic is to determine how LLMs can be utilized for cybercrime and systematically analyse the benefits for malicious actors.

6 Security of LLM Generated Source Code

Schreiber and Tippe [1] analyzed in their recent work the influence of LLMs on security of real software projects. In their study, they extracted LLM-generated code from GitHub via keywords and than used static code analysis to detect security issues.

LLM-Code-Sec: Building upon this work, the methodology should be extended to also discover code that was likely generated by LLMs without explicit attribution. Instead of using standard software tests, advanced methods should be utilized to detect real security issues rather than checking for static code patterns.

7 Deep Fake Media: Statistical Realism and Visual Fingerprinting in Synthetic Media

As generative models reach a state of near-visual perfection, the distinction between sensor-captured imagery and synthetically generated content has moved from the visible domain to the latent and statistical domains. Current research indicates that AI-generated images possess unique mathematical fingerprints due to the specific optimization goals of their underlying architectures [2]. For instance, generative models often exhibit uneven color distributions and vibrant idealized peaks that differ from the balanced distributions produced by the color correction matrices and white balance algorithms in modern smartphones [2]. Furthermore, recent investigations into pre-trained feature spaces, such as CLIP (Contrastive Language-Image Pre-training), reveal that real and synthetic images occupy distinct topological regions within the latent manifold [3]. These discrepancies are often a byproduct of perceptual loss functions (e.g., LPIPS) that prioritize luminosity over chrominance, leaving subtle residuals in the chroma channels that are invisible to the human eye but identifiable via forensic analysis [4].

The following topic explores the convergence of these two domains, focusing on how image statistics can be adapted to bridge the gap between biological perception and algorithmic detection.

IMG1: This topic explores the statistical properties of real images compared to synthetic ones and the state of current detection algorithms.

8 Forensics and Anti-Forensics

The field of forensics is constantly exposed to new technologies and software versions that all leave different traces. These traces are the foundation for investigators to solve criminal cases and in the final instance secure a conviction in court.

FOR1: The objective of this topic is to conduct a systematic analysis of currently available anti-forensic methodologies and tools, evaluating their impact on forensic workflows.

FOR2: If you already have working experience in the digital forensics fields, it might be possible to generate an individualized topic together. Topics can range from mobile phone forensics to network forensics, provided they maintain a strong research focus (rather than purely practical application). Previous work from the DFRWS conference[1] assist in getting an impression of suitable topics. The research topic must address new questions or unexplored/underexplored aspects of existing problems and must fill identified gaps in current knowledge.

9 Cryptocurrency and Privacy

In recent years, financial sanctions have emerged as a prominent policy instrument in the digital asset space (cryptocurrencies). Regulatory bodies, notably the U.S. Department of the Treasury's Office of Foreign Assets Control (OFAC), have demonstrated an increasing willingness to target cryptocurrency entities, individuals, and critical infrastructure components perceived to facilitate illicit finance or sanctions evasion. The academic community is actively working to understand the effectiveness of these crypto-specific sanctions. Early empirical studies, such as the work by Zola, Medina, and Orduna [5] present a nuanced picture, indicating mixed results: while some sanctioned entities appear to be deterred, a significant portion continues their operations, often adapting their tactics to circumvent restrictions. These findings suggest that sanctions are not a simple *off switch* for illicit crypto activities. Effective sanctions must be more complex than blacklisting bitcoin addresses.

BIT1: Can novel heuristics reliably detect the use of specific advanced PETs when applied to Bitcoin transactions or Bitcoin-derived assets on other chains? This includes identifying subtle on-chain signatures of i.e. ZKP constructions, or characteristic interaction patterns with DeFi protocols that are indicative of layering or obfuscation. For instance, could behavioral or on-chain metadata heuristics be devised for Bitcoin transactions that interact with cross-chain bridges or exhibit patterns consistent with known state-sponsored APT laundering methodologies? A starting source could be [6].

BIT2: This research direction involves the design, implementation (as a Proof-of-Concept, PoC), and forensic analysis of a Bitcoin mixer. The goal is to explore critical design decisions that mixer developers face and to evaluate how these decisions impact the mixer's privacy guarantees and its susceptibility to various forensic analysis techniques. This constructive approach allows for a deep, practical understanding of mixer vulnerabilities and strengths from both a user's and an investigator's perspective.

BIT3: TRON (TRX) has emerged as a critical settlement layer, particularly for the Tether (USDT-TRC20) stablecoin, with transfer daily volumes exceeding $600 million. Despite this scale, systematic forensic research on TRON remains remarkably scarce compared to Bitcoin and Ethereum. This research track explores the identification of illicit infrastructure within the TRON ecosystem, specifically focusing on the role of *Resource Pledging* (Energy/Bandwidth delegation) and the *Memo* field as metadata vectors for laundering coordination.

The investigation targets the mapping of shadow banking networks, where criminals utilize the low-fee TRON environment to perform high-velocity layering. Key research questions could include:

[1] https://dfrws.org/presentation/

1. Can the analysis of resource feeder wallets (which delegate Energy to thousands of disposable mule accounts) unmask illicit clusters that are otherwise disconnected in value-transfer graphs?
2. To what extent does the TRON *Memo* field (transaction metadata) serve as an on-chain coordination layer?
3. How can cross-chain analysis be optimized to link illicit funds to their eventual exit via TRON-based USDT? Initial frameworks for large-scale data extraction from TRON, such as the work by [7], provide a foundational starting point for this forensic exploration.

References

[1] Maximilian Schreiber and Pascal Tippe. "Security Vulnerabilities in AI-Generated Code: A Large-Scale Analysis of Public GitHub Repositories". In: *Information and Communications Security*. Springer Nature Singapore, Oct. 2025, pp. 153–172. ISBN: 9789819535378. DOI: 10.1007/978-981-95-3537-8_9. URL: http://dx.doi.org/10.1007/978-981-95-3537-8_9.

[2] Zexi Jia et al. "Secret Lies in Color: Enhancing AI-Generated Images Detection with Color Distribution Analysis". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2025.

[3] Lea Uhlenbrock, Sandra Bergmann, and Christian Riess. "Latent Landscapes: Topology of the CLIP Feature Space for Synthetic Image Detection". In: *European Signal Processing Conference (EUSIPCO)*. 2025.

[4] Lea Uhlenbrock, Davide Cozzolino, et al. "Did You Note My Palette? Unveiling Synthetic Images Through Color Statistics". In: *Proceedings of the 2024 ACM Workshop on Information Hiding and Multimedia Security (IHMMSec)*. 2024.

[5] Francesco Zola, Jon Ander Medina, and Raul Orduna. *Assessing the Impact of Sanctions in the Crypto Ecosystem: Effective Measures or Ineffective Deterrents?* 2024. arXiv: 2409.10031 [cs.CR].

[6] Jan Zavřel et al. "Tumbling down the stairs: Exploiting a tumbler's attempt to hide with ordinary-looking transactions using wallet fingerprinting". In: *Forensic Science International: Digital Investigation* 52 (2025). DFRWS EU 2025 - Selected Papers from the 12th Annual Digital Forensics Research Conference Europe, p. 301869. ISSN: 2666-2817. DOI: https://doi.org/10.1016/j.fsidi.2025.301869.

[7] Qian'ang Mao et al. *Decoding TRON: A Comprehensive Framework for Large-Scale Blockchain Data Extraction and Exploration.* 2025. arXiv: 2509.16292 [cs.CR]. URL: https://arxiv.org/abs/2509.16292.