SE assignment

Antonio Napoli

August 2025

1 Introduction

My research project presents a novel approach to determine when two machine states are observationally equivalent by synthesizing a projection function (Observation) using the Syntax-Guided Synthesis (SyGuS) methodology and cvc5 in order to create a leakage model aggregating multiple observations in many iterations.
We explore the problem of observational equivalence in machine states consisting of 64-bit vector registers, with examples initially derived from a ground truth model representing the ideal leakage.
The goal is to improve the efficiency of the synthesized projection function, which distinguishes equivalent and non-equivalent states after executing specific instructions. Our main contributions are : optimization of the synthesis with the help of large language models (LLMs), by extracting constants and subexpressions from CPU technical documentation to augment SyGuS grammars; refinement of the observation synthetized, generating example that invalidate the observation with respect to the ground truth; reuse part of the observation found as subexpression to enrich the grammar. This enables enumeration-based synthesis that avoids SMT-based constant generation and optimizes the process. Queries are instruction-dependant and CPU-dependent, with the LLM providing constant values and sub-expression that reveal operand-dependent leakage (e.g., ARM7TDMI multiply instruction) or memory-dependant leakage (Tag-idx). The area touched by this project are multiple : Usage of LLM, software side-channel detection, verification, SMT solving and optimization.
The main focus is put in the optimization area, since the goal aim to detect leakage previously undetected due to high computational cost.

The usage of LLM is mainly from a user application side, where we use tecnique as RAG, reasoning and task splitting in order to reduce the search-space of the SAT solver, crafting a CPU tailored grammar.
The security of side-channel while it's the main reason that led to this work, it's somehow marginal and we reuse methods and definition previously presented in [].

2 Lecture principles

One of the main point observed in the lecture it regards the quality and assurance in SE and ML.
The end goal of the project is to reduce complexity with an ML approacj and increase efficiency, as mentioned: build the product right; but since it's a verification process, and it involves the detection of an information leakage, using a sat solving tool, we ensure that our output is reliable to respect of the data taken as input, verifiable and reporducible.
We add an ulterior task at the end of the process with multiple test, where the data are SAT generated, to measure the soundness and the precision of our formula, in order to check if we overapproximated or underapproximated in our output.
This usage of the LLM encoded in out project also display the need of ML engineering in the building process. Even if the system used is OpenAI API, and the implementation is rather easy compared to build an entire Neural Netowrk or other complext ML tecniques, we noticed that in order to not slow down the evolution of the project on the science side it was needed to build a reliable integration.
With the change of the idea -scientifically talking- it was needed to restructure the whole implementation since it was very hard scoped, furthermore we decided to create a standalone tool for this in order to use it and integrate in other process used by other researcher.

3 Guest Lectures

Julian Frattini, Ph.D, emphasized in his lecture Julian Frattini, Ph.D.the importance of requirement engineering.

This phase even if critical it was not considered at the start of the project, and since the multiple parts involved in the tasks, this lack of requirement handling became a problem later on.
I would argue that usually it might be difficult to write down specific technical requirement in the solution space, since PhD projects have to produce innovation and usually are goal oriented and define in the problem space.
It is difficult to know the technical requirement in order to reach the goal, nevertheless trying to sketch a possible idea would surely help and avoid further deviation that might mislead later on.

4 Data Scientists versus Software Engineers

In the book "Machine Learning in Production", it's described the difference by the author between Data Scientist and Software Engineers.
I broadly agree with this difference, because Data Science usually they are mostly focused on the Model-Algorithm scope, and they overlook on other corner technicalities such a depolyment, integration, robustness and so on.
Usually these latest are handled by Software Engineer, which, on the other hand, might lack experience in optimizing the performance of a model, while still being able to fully deploy and utilize one.
These two roles might converge in the future, and I see more potential in the Data Scientist that might make up their lack of technical programming skill also thanks to the advent of the LLM, which might be able to fully deploy a solution with the right indications given by a Data Scientist.
On the other hand I can't see the same takeover from the Engineers, because the theoretical knowledge behind the ML is crucial and I doubt LLM might ever help in optimizing a "new problem".

5 Paper analysis : LLMs for Test Input Generation for Semantic Caches

5.1 Core Ideas and importance to the SE

The core idea of this paper is to introduce a technique to mitigate some errors in LLM system such as false positive caused by the cache reuse.

5.2 Relation to your research

The relation to my research is in the test input generation aspect. The quality of the inputs directly affects the validity of the synthesized function or observation. Their idea of using LLMs to generate subtle variations of queries it might help with my own use of LLMs to generate constants, subexpressions, and instruction-specific variations for SyGuS grammars. .

5.3 Integration into a larger AI-intensive project

In a larger AI-intensive project such as a Secure CPU Verification Platform, semantic caches might be used for efficiently serving verification queries across distributed nodes. The VaryGen approach could be applied to automatically generate workloads of queries that stress-test the cache before deployment.

5.4 Adaptation of your research

The knowledge gained fromm this paper might help in the future in my own research.
I will first have to address if this problem persists also in my query and I will evaluate the impact.
Since a main part of my project is LLM dependant, and the area is security-formal verification, any improvement would make a difference.
Indeed the structure of my work and this are different, such as the main goal and the dataset used, but they have many points in common that could be abstracted and reused into my own research.

6 Second Paper : Bringing Machine Learning Models Beyond the Experimental Stage with Explainable AI

6.1 Core Ideas and importance to the SE

This paper presents an analysis done in a Danish Bank, where a ML model was successfully integrated into the day-to-day work and it presents the main reason which made it possible.

6.2 Relation to your research

The relation is not immediate to my research, but in the broader area of software security.
I'm mainly interested in the Explainable AI because in software security it's not enough to know if a piece of code, or an architecture, is vulnerable or not, it's needed to understand what causes the vulnerability.

6.3 Adaptation of your research

Ad example into my pipeline I do start with an input of data labeled as : information leakage, sound.
And the goal of my project is to understand what the cause behind this separation.
In a possible future research, I plan on working on using an ML model with explainability.
In this way I would cancel the need of a SAT-solver, or at least make it secondary, and focus on performance given by these models.
At the state of the art, the main limitation is based on the performance since the generation of more datapoint is negative for a solver, but on the other hand ML is not usable because it would miss the reparation idea.
The explainability would be a good idea that would fit both problems.

7 Research Ethics and Synthesis Reflection

I surfed the CAIN main website, and looked into the latest 3 years of paper published.
From here I mainly opened papers in the area of LLM and Explainable AI, since these were the topic I was interested in.
In order to ensure originality I did use only the material mentioned by the assignment file, and the papers were found directly on the CAIN website.
I wrote without the use of LLM or external tools.