LLM-Augmented Static Analysis Security Testing

Murilo Escher Pagotto Ronchi

Seminar: Software Quality
Advisor: Kohei Dozono
Technical University of Munich
murilo.escher@tum.de

Abstract. The abstract should briefly summarize the contents of the paper in 15–250 words.

Keywords: First keyword · Second keyword · Another keyword.

1 Plan

The idea of this project is to use RAG, consisting of CVE reports and SAST tools, to enhance a LLM's ability to perform vulnerability detection. The focus is going to be on reducing false positives for memory related vulnerabilites.

Using this dataset https://github.com/ARISE-Handong/BugOss, the pipeline will be as follows:

1. Run CodeQL on dataset
2. Get related CVE reports based on the output from CodeQL
3. Merge both information as RAG ressource
4. Use Gemini's API to send this info, along with the vulnerable code to the LLM
5. Get LLM response through API
6. Save response in a table?

The data then will be evaluated as follows:

1. The false positive rate will be calculated for both CodeQL and the LLM
2. The #vuln detected must be taken into account, otherwise FP would be 0 for 0 vulns

Would using another SAST tool with a voting system (both must detect the vuln. for it to be sent to the llm) be useful?

2 Introduction

Large Language Models (LLMs) have recently gained a lot of popularity. Among their many uses, one currently much researched possibility is their application in Static Application Security Testing (SAST).

Talk about repo-level being more important, [1]

In this paper, we further investigate their usability when combined with SAST Tools.

3 Study Design

How the tools were chosen, flow of information, etc

4 Evaluation Setup

How the results are going to be interpreted. For ex. the voting system when using multiple SAST Tools. The research questions we are going to analyse also come here.

5 Results

Here I am going to discuss the results obtained, use tables or any other way to show info and compare. There should be a clear discussion on how the methods improved the baseline, etc.

6 Related Work

Here it should be made clear how this project differs from related work. The ways in which we improved or completed other research should be made clearer.

7 Conclusion

You can also reference other parts of the document, e.g., sections or subsections. In Section 2 we briefly introduced something, whereas in Subsection **??**, we motivated something else.

Make sure to capitalize chapters, sections or subsections when referencing them.

A Appendix

Anything additional goes here . . .

Maybe extra information as to how the code words could be provided, if needed.

References

1. Risse, N., Böhme, M.: Top score on the wrong exam: On benchmarking in machine learning for vulnerability detection (2024), `https://arxiv.org/abs/2408.12986`