Department of Computer Science
Amrita School of Computing, Amritapuri Campus
B Tech CSE (2022 Admission)

19CSE499 PROJECT PHASE-II

PROJECT PROPOSAL SUBMISSION

Group Number: C - 3

Team Members:

Roll No	Student Name
AM.EN.U4CSE22002	ADABALA NAVEEEN SAI KUMAR
AM.EN.U4CSE22223	GOLUKONDA RAGHUVEER
AM.EN.U4CSE22230	CHAPPIDI JASWANTH
AM.EN.U4CSE22260	YERUBANDI HARSHITH

Project Guide: Aswathy Mohan

Project Title : Intelligent Bug Detection and Code Clone Identification Using Hybrid Analysis Techniques

Novelty :

The proposed project brings forth an integrated hybrid approach that amalgamates static code analysis, semantic understanding by an LLM, and light execution analysis for concomitant removal of bugs and identification of code clones simultaneously. The proposed approach seamlessly amalgamates all the above-mentioned factors that standard tools alone cannot pick on while identifying semantic clones and bugs at runtime effectively.

Objectives :

Overall Objectives:

- Development of an intelligent system that can detect both bugs and code clones in the software with the help of a hybrid analysis process.
- To deal with the drawbacks of static-only and dynamic-only approaches by incorporating structural, semantic, and behavioral information.
- To help software developers enhance the quality and dependability of software in larger software systems.

Phase-II (S8) Objectives:

- Create AST-based static analysis tools to identify structural patterns from Java and Python source codes.
- Apply the utilization of semantic embeddings provided by the LLM to detect strong Type-3 and Type-4 semantic clones.
- Develop a lightweight component for dynamic analysis through the use of fuzzing and bytecode instrumentation.
- Create a fusion engine that combines the static, semantic, and dynamic scores for enhanced detection.
- Generate clone and bug reports that should be understandable by humans, determining the severity levels
- Testing on open repositories and open datasets.

Novelty - Objective mapping :

The novelty of the hybrid multi-modal analysis comes from relating each objective to a different kind of detection signal. The structural well-foundedness comes through static analysis, semantic embeddings represent the intent of logic, while dynamic execution checks semantics of real behavior. Combining these objectives helps novelty in the project by allowing semantic clones & hidden bugs which are undetectable through single objectives.

Scope of the project (S8) :

Inclusions (What is covered):

- Static analysis based on AST and control-flow.
- Semantic analysis via transformer embeddings of code.
- Lightweight Dynamic Analysis by Fuzzing and Java Bytecode Instrumentation.
- Type-1, Type-2, Type-3, Type-4 code clones
- Runtime bug identification, as well as bug propagation caused by clone development.
- Generation of reports about bugs and code clones.

Exclusions (What is NOT covered):

- Full symbolic execution or exhaustive path search.
- Autosuggestions for fixing or refactoring code.
- GPU-intensive Model training.
- Cross-platform IDE plugin deployment.

Deliverables:

- A functioning prototype of the hybrid bug and clone detection system.
- Results of the evaluation of the experiments.
- Documentation of system architecture and workflow.

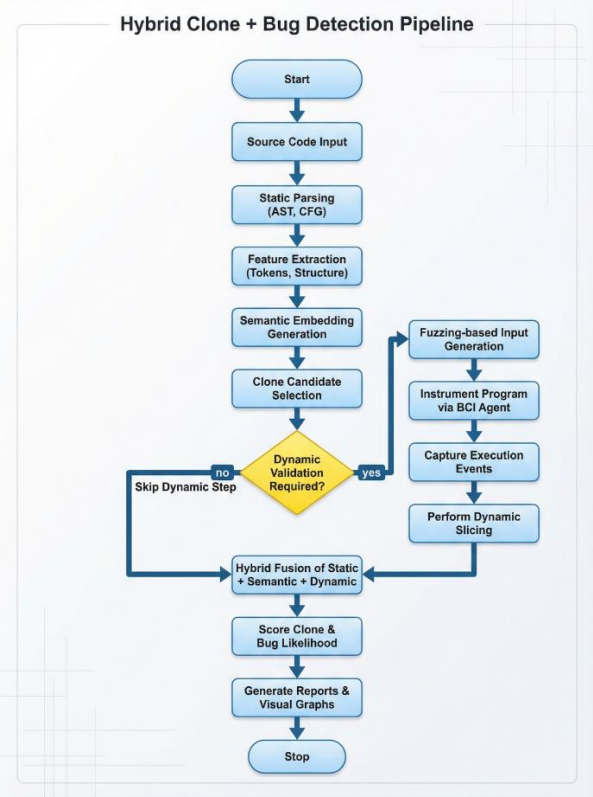
Boundaries & Limitations:

- Coverage in dynamic analysis depends on the inputs and paths that are generated.
- Semantic embeddings are based on pre-trained models, which can overlook very strongly domain-specific logic.
- The current supported programming languages by this system are Python and Java.

Expected Outcomes:

- Correct identification of semantic code clones.
- Coverage of hidden bugs that are missed by static analysis tools.
- False positives reduced via multi-signal validation.
- Enhanced understanding of code quality and related risks for developers.

Project Design:



Design- Objective mapping :

The layered architecture is directly helpful to Phase-II goals as it ensures that every goal is achieved as a separate yet internationally operable module. Static analysis is helpful in achieving structural detection tasks, semantic embeddings facilitate deep clone detection, dynamic analysis verifies correctness at runtime, and the fusion engine combines every signal to obtain maximum precision as well as recall.

Signature of the guide: