Dynamic, Multi-Language Verification of GDPR Compliance in Scripts: From Prototype to Evaluable Tool

- Advisor: Alain Tchana

Description

Research scripts often collect personal data (cookies, browser fingerprints, system identifiers), which makes them subject to the GDPR. Today, validating them is a slow, manual process: in one real-world case, the distribution of a cookie-analysis script was halted until a lengthy review by compliance experts with limited familiarity with code had been completed. Existing tools (RuleKeeper, Sesame, Paralegal) target web applications written in a single language (JavaScript or Rust) and rely on static analysis.

A previous internship proposed a different approach. A manifest readable by non-programmers describes a script's sources, sinks, data categories, purposes, and allowed actions. A Truffle/GraalVM-based analyzer dynamically instruments the script's AST to verify that it complies with this manifest, regardless of the language it is written in.

A JavaScript prototype is already functional.

Objectives

- Extend the analyzer to a second language (Python).
- Extend the analyzer to AI agents.
- Translate several representative scripts (cookie collection, fingerprinting, system measurements) into manifests to assess the expressiveness of the DSL.
- Measure detection on intentionally faulty scripts: data leaks, undeclared sources, disallowed actions.
- Measure runtime overhead.