Software Design and Development Engineer (M/F): vulnerabilities in code generated by LLMs (TAP project)

Contract type: Fixed-term contract
Start date: As soon as possible
Duration: 13 months, extendable to 35 to 40 months (depending on experience)
Required degree level: Master's degree or equivalent

Scientific and Technical Context

The TAP Project (Trustworthy Automatic Programming)

TAP is a joint project between CNRS-IPAL-IRISA and NUS (National University of Singapore), funded by the DGA and its Singaporean counterpart.

Over the last 60 to 70 years, programming has dominated computer science, involving the capture of intentions and code production. Formal specifications have gained prominence thanks to advances in system modeling and design, allowing for more precise goal capture. Despite these advancements, software engineers are often reluctant to write formal specifications. This leads to the absence of formal declarations of intent for large software systems, making debugging and error correction difficult. In the absence of formal intent capture, testing and analysis have been used to develop reliable codebases. Testing aims to achieve broader behavioral coverage, employing test oracles. Fuzzing approaches have become significant over the past decade. However, achieving functional correctness of software without in-depth formal requirements remains a challenging goal.

Recent advances in automatic code generation from large language models (LLMs) provide a new perspective. It is now conceivable to program based on natural language specifications using LLM code generation, suggesting that auto-coding is feasible. This raises the issue of the correctness and security of code automatically generated by LLMs and the conditions under which this code can be trusted.

The **TAP Project** focuses specifically on these aspects. The project aims to identify vulnerabilities in LLM-generated code, analyze and classify these vulnerabilities, and determine if certain types of vulnerabilities are more common in LLM-generated code compared to human-written code. The project also seeks to automate the correction of these vulnerabilities and improve LLMs concerning code vulnerabilities.

Tasks and Responsibilities

The primary goal of the DiverSE team in this project is to identify vulnerabilities in LLM-generated code. To achieve this, we will develop a system capable of automatically generating datasets of vulnerabilities. This will be done using available web catalogs of vulnerabilities and modeling these vulnerabilities for seamless integration into a testing tool, enabling the analysis of LLM-generated code and libraries. The target programming languages will primarily be **C** and **Java**, given their widespread use and to maximize the impact of our work.

In this context, the DiverSE team (in close collaboration with the IPAL laboratory and the DGA) is recruiting a software design and development engineer for **13 months**, extendable to **35 to 40 months** (depending on experience). This role will be under the scientific and technical supervision of DiverSE team members involved in the project. The engineer will be responsible for the design and development tasks related to DiverSE's objectives, including the creation, implementation, and presentation of prototypes, demonstrators, and datasets. Synergies with other ongoing projects within the team will also be explored and leveraged. The project results will be used by our NUS partners in Singapore.

The position may involve travel in France and abroad, including air travel.

Skills

- **Software design and development skills** required; experience is a plus.
- Knowledge of **C** and **Java** is a plus. Generally, candidates are expected to be proficient in multiple programming languages.
- Ability to work in an international environment and communicate in **English** is appreciated.
- A good level of **autonomy** is valued.

Work Environment

IRISA (Research Institute of Computer Science and Random Systems) is one of France's largest research laboratories in computer science and information technology, with over 850 members. Organized into seven scientific departments, IRISA focuses on key areas such as bioinformatics, system security, software architectures, virtual reality, big data analysis, and artificial intelligence.

IRISA is part of a dynamic regional ecosystem, recognized for its expertise through international scientific collaborations. Focused on the future of computer science, IRISA plays a key role in

digital transformation, cybersecurity, health, environment, transportation, robotics, energy, and AI.

The **DiverSE research team** specializes in software engineering techniques for building reliable and efficient applications, focusing on areas such as cybersecurity and LLMs. The team consists of about 15 permanent members (Inria and CNRS researchers, university lecturers, including 3 members of the French University Institute), 15 PhD students, several engineers, and a DGA associate engineer. DiverSE is internationally recognized and maintains strong ties with global, national, and local industries. The team also prides itself on a friendly and engaging work atmosphere.

The position is located in a sector covered by the protection of scientific and technical potential (PPST), and therefore requires, in accordance with the regulations, that your arrival be authorised by the competent authority of the MESR.

Why Join Us?

Project Highlights:
This project offers unique opportunities due to its application domain, ambition, international network, and potential impact. It lies at the core of DiverSE's activities and involves collaboration with a dynamic team in Singapore.

Ambition:
You will contribute to a worldwide open-source project. In an era where source code security is a strategic concern, TAP aims to address this challenge directly. This project could also lay the groundwork for stronger collaboration between DiverSE and NUS, enhancing national, European, and global sovereignty and security in software engineering, AI, LLMs, and cybersecurity.

Network:
TAP involves frequent interactions with partners from NUS and IPAL. Visits to Singapore may be arranged based on your preferences. The project offers opportunities to engage with various research, innovation, and industry transfer projects within and beyond the DiverSE team. After the project, you'll be one of the (many) alumni of the DiverSE team, most of whom are still in touch.

Impact:
The exponential growth of LLM usage for code generation ensures significant impact potential. Automating the securing of LLM-generated code addresses a pressing global need, with substantial cybersecurity implications.

Benefits

- Remote work up to 2 days per week.

- Partial reimbursement of public transport or sustainable mobility costs.
- Partial coverage of health insurance costs.
- Subsidized on-site dining.
- Free car and bicycle parking; bus stop 5 minutes away; metro station 10 minutes away.

Salary

Monthly salary based on degree and experience:

- **€2,847 gross** (€2,288 net) to **€3,514 gross** (€2,824 net).

Location

Campus de Beaulieu, IRISA/Inria Rennes
Building 12
263 Avenue du Général Leclerc
35042 RENNES Cedex, France

Contacts

- **Olivier BARAIS**, Professor, University of Rennes: Olivier.Barais@irisa.fr
- **Olivier ZENDRA**, Researcher, Inria: Olivier.Zendra@inria.fr