Using machine-learning and NLP to analyze open-source software repositories

Opportunity for a 6-month internship

Security Research @ SAP Labs France
Sophia-Antipolis – France

Maintaining security is a constantly shifting task, and we need to respond with continuous learning and research. The portfolio of SAP Security Research contains those topics that we believe are most important for SAP's security future.

SAP's vision to secure business is built on 3 ideals: **Zero-Vulnerability**, to harden the software by eliminating vulnerabilities, **Defensible Application**, to enable the software to identify and prevent attacks, and **Zero-Knowledge**, to make any theft of data useless through encryption.

Considering these aspects, SAP Security Research covers the following focal areas: Anonymization for Big Data, Secure Internet of Things, Software security analysis, Open-source analysis, Deceptive application, Applied cryptography, Quantum technology, and Machine Learning as enabler for the next generation of security.

Security Research proposes a 6-month internship in its Sophia-Antipolis offices (Mougins, France).

INTERNSHIP TOPIC

Today, tools supporting such impact assessments rely on so-called vulnerability databases such as the NVD, which are enumerations of known software vulnerabilities. Those databases, however, cannot provide complete coverage, i.e., many known vulnerabilities will never be listed.

To reduce the dependency on these sources, SAP Security Research investigates novel approaches, leveraging machine-learning (ML) and methods originated in the natural language processing (NLP) field to analyze source code repositories and to automatically identify commits that are security-relevant (i.e., that are likely to fix a vulnerability, or that introduce a new vulnerability) [1].

While the current results are encouraging, SAP Security Research now focuses on improving the predictive performance to obtain more accurate predictions and to scale to real-life scenarios. In particular, our team is working on defining (or automatically learning) better features, on how to efficiently extend the size of annotated resources at our disposal, and in particular on how to combine different textual resources (commits, pull requests, mailing list discussions, bug-tracking tickets, security advisories, etc.) to gather more information which the prediction can be based upon.

This internship aims at developing a method to automatically map security advisories onto the source code commits that address and mitigate them. To devise such method, the student will explore the application of different techniques involving cutting-edge machine learning models and natural-language processing.

We expect that 40% of time will be dedicated to research activities, and 60% to development.

[1] A. Sabetta, M. Bezzi, "A Practical Approach to the Automatic Classification of Security-Relevant Commits", 2018. Available online: arxiv.org/abs/1807.02458

CANDIDATE PROFILE

- University Level: Last year of MSc or less if the student has a good profile
- Solid foundations in CS and a passion for well-designed, cleanly implemented software
- Good knowledge of one or more of the following languages: Java, Python
- Experience with GIT, Linux (bash)
- Knowledge in (or interest in learning) machine learning basics
- Knowledge of one or more of the following is desirable (but not required): pandas, spacy, scikit-learn, keras, tensorflow.
- Interest in experimental research
- Fluency in English (working language)
- Good oral and written communication skills

INTERNSHIP CONTEXT

SAP

Founded in 1972, SAP has grown to become the world's leading provider of business software solutions. SAP is market leader in enterprise application software. The company is also the fastest-growing major database company. Globally, more than 77% of all business transactions worldwide touch an SAP software system. With more than 347.000 customers in more than 180 countries, SAP includes subsidiaries in all major countries. SAP is the world's largest inter-enterprise software company and the world's third-largest independent software supplier, overall. SAP solutions help enterprises of all sizes around the world to improve customer relationships, enhance partner collaboration and create efficiencies across their supply chains and business operations. SAP employs more than 98.600 people.

Security Research at SAP Labs France, Sophia Antipolis

Based at SAP Labs France Mougins, Security Research Sophia-Antipolis addresses the upcoming security needs, focusing on increased automation of the security life cycle and on providing innovative solutions for the security challenges in networked businesses, including cloud, services and mobile.

STANDARD INTERNSHIP PACKAGE

- *Salary*: depending on the length of the internship and your diploma.
- *Lunch*: SAP Labs France has a local cafeteria; interns contribute 2,63 €uro/lunch, like other SAP employees.
- *Holidays*: French Bank Holidays
  - January 1st; April 12th, April 13th, May 1st, May 8th, May 21st, June 1st, July 14th; August 15th, Nov 1st and 11th; December 25th
- *Travel*: no trip will be paid by SAP.
- *Accommodation*: SAP can propose an accommodation for the duration of your internship. The accommodation is subsidized by SAP: the intern pays half of the rental cost: 342€ for a 1-room apartment or 442€ for a 2-room apartment (Choice depending on the availability).

CONTACTS AND PROCEDURE

Please candidate by clicking on this link:
https://career5.successfactors.eu/sfcareer/jobreqcareer?jobId=234529&company=SAP&username=

UPLOAD (all documents must be in English):
- Your CV
- Cover letter
- Any relevant documents