Internship+Ph.D. Offer — SDS Team — INSA CVL

Towards Formal Semantics and Proven Compliance of Business Workflows Processing Personal Data

Contact: Vincent Hugot — vincent.hugot@insa-cvl.fr

You can check that you have the latest version of this document and that the position is still available on https://tcs.vhugot.com/Offers.

1 Practical Information

Laboratory & Team:

Laboratoire d'Informatique Fondamentale d'Orléans (LIFO, EA 4022),
Systems and Data Security (SDS) team,
INSA Centre Val de Loire,
88 boulevard Lahitolle
18022 Bourges

Supervisors and collaborators:

The internship will be supervised by Vincent Hugot.

The following Ph.D. thesis will be co-supervised by Vincent Hugot and directed by either Pascal Berthomé or Benjamin Nguyen, depending on scope.

The workflow software company Relyens/QualNet[a] shall also offer support in the form of real-world workflows and use-cases, as well as access to their IntraQualDyn software.

Duration and start:

The internship will last 5 to 6 months, and start at the candidate's earliest convenience, sometime around March, April, or May 2024.

[a] https://www.qualnet.fr/

It serves as prelude to, and groundwork for, a Ph.D. thesis, which shall be funded (3 years) by the iPoP project[b]. Do not apply if you have no intention of pursuing a Ph.D. after.

Candidates with a Master's degree and an excellent background may also apply directly to the Ph.D. position. The starting date is more flexible in that case.

Contact:

To apply or request additional information, send an email to `vincent.hugot@insa-cvl.fr`, preferably with a `[Workflows]` prefix to the mail title.

To apply, please enclose:

- your CV,
- your academic transcript for your Bachelor and Master's degrees (or equivalent),
- a short cover letter explaining your motivation for pursuing a Ph.D., and how you meet the requirements and skills below,
- the contact information of one or two referees; specify in what capacity they know you.

Before applying, please check that you have the latest version of this document and that the position is still available: `https://tcs.vhugot.com/Offers`

Requirements:

- For the internship: Being in the final year of a Master's degree in computer science, computer engineering, mathematics, or equivalent.
- For the Ph.D.: Holding a Master's degree, as above.
- Proficiency in written English. Fluency in spoken English or French.
- Exposure to, and interest in, automata theory, formal grammars, formal verification, theorem proving, and logic.
- Good programming skills (emphasis on Python).

Useful Skills and Traits:

- Previous exposure to modal logics, CTL, LTL, CTL*, Büchi Automata, Model-Checking, etc, or theorem-proving software such as Coq, Isabelle/HOL, etc
- Though we shall work in Python, previous exposure to, and taste for, functional programming, in particular OCaml, Haskell, etc, would probably smooth things out.
- A willingness to navigate documents related to private data handling, such as the GDPR.

[b] `https://files.inria.fr/ipop/`

2 Context and Long-Term Goals

This project builds on a previous collaboration between SDS and Qualnet, during which we defined a formal semantics for (a subset of) the Intraqual graphical business process workflow language.

Note that proper formal semantics are very uncommon for business workflow languages, which tend to be defined graphically and given informal semantics, usually in natural language.

The short-term aim was to produce a principled, reliable execution engine and ensure correctness of execution. We are now in a position to apply formal verification techniques to workflows.

Of particular interest for our project are workflows dealing with personal data generally, and sensitive personal data especially; for instance, workflows in the medical domain.

During the Ph.D., we hope to develop methodologies to prove that workflows are in compliance with regulations on handling of private data. Thus we shall:

- ◇ continue developing workflow languages with formal semantics;
- ◇ identify which features and design principles of workflow languages are compatible or incompatible with formal verification and proof techniques;
- ◇ identify provable/verifiable classes of properties for those languages, with a special interest for properties applicable to data handling. Generally, this means controlling who has access to the data, or to specific *views* of the data (anonymised, differentially private, etc);
- ◇ study to what extent proven properties of the workflows can be transferred to automatically generated websites or software implementing the workflows;[c]
- ◇ study to what extent legal requirements regarding data handling can be formalised and proven on our workflows and compiled websites;
- ◇ test the practical robustness of our solution on real-world workflows and public data.

The aim is not an *immediate* industrial application; we shall stay agnostic with respect to specific languages and technologies. All research shall be publicly available, and all code open-sourced. However, QualNet will provide us with real-world workflows and use-cases, which should guide our modelling work.

[c] The IntraQualDyn software, for instance, automatically generates a website that implements the workflow, providing each user with the data and data entry forms they need at the current step in the process.

Correctness of the workflow on paper does not by itself imply correctness of the software; that requires additional work ensuring that the “compiling” process from workflow to software preserves the proven properties

3 Short-Term Goals of the Internship

The main goal of the internship is to lay the groundwork for the Ph.D. topic, as above. In particular, some time should be spent on exploring the prior art on workflow semantics and proof, and some time on reading real-world workflows and confronting them to the GDPR and similar documents to identify interesting properties to verify. This will inform the starting directions of the Ph.D.

A secondary project, the implementation of a CTL* model-checker (in Python), will be pursued in parallel. It is a more self-contained subject, and serves mainly as a pretext to test and practice the skills that will be necessary for completion of the Ph.D.