The CoLiS project aims at applying formal verification techniques to the analysis of Shell scripts, in particular those that are used for installation of software packages. These scripts play a central role in the installation of computers under UNIX and Linux. They interact with the state of the operating system, in particular at the level of the file system. Mistakes in such scripts may have damaging impacts since they are executed at critical moments, in administrator mode, in multiple situations and scenarios.

Another initial step of the CoLiS project is the implementation of a parser of the POSIX shell language, in order to analyze the real installation scripts. This syntactic analyzer already allowed us to compute statistical data on the full Debian distribution to figure out which constructs were widely used and which were not. A translation tool of these scripts into the CoLiS language is in progress.

The initial analysis of the needs lead us to decide that a useful method for analyzing scripts semantics in a fully automated way is symbolic execution. The first goal of this post-doc position is to participate to the design of this symbolic interpreter, in a formal way using Why3, building up on the already formalized concrete interpreter. A formal proof of the correctness of the symbolic execution with respect to the concrete semantics is expected. Also, a symbolic interpreter is tightly connected with a constraint language to describe symbolic states, in that case it will be a constraint language of the file system tree.

Another initial step of the CoLiS project is the implementation of a parser of the POSIX shell language, in order to analyze the real installation scripts. This syntactic analyzer already allowed us to compute statistical data on the full Debian distribution to figure out which constructs were widely used and which were not. A translation tool of these scripts into the CoLiS language is in progress.

The CoLiS project aims to apply formal verification techniques to the analysis of Shell scripts, in particular those that are used for installation of software packages. These scripts play a central role in the installation of computers under UNIX and Linux. They interact with the state of the operating system, in particular at the level of the file system. Mistakes in such scripts may have damaging impacts since they are executed at critical moments, in administrator mode, in multiple situations and scenarios.

Another initial step of the CoLiS project is the implementation of a parser of the POSIX shell language, in order to analyze the real installation scripts. This syntactic analyzer already allowed us to compute statistical data on the full Debian distribution to figure out which constructs were widely used and which were not. A translation tool of these scripts into the CoLiS language is in progress.

The initial analysis of the needs lead us to decide that a useful method for analyzing scripts semantics in a fully automated way is symbolic execution. The first goal of this post-doc position is to participate to the design of this symbolic interpreter, in a formal way using Why3, building up on the already formalized concrete interpreter. A formal proof of the correctness of the symbolic execution with respect to the concrete semantics is expected. Also, a symbolic interpreter is tightly connected with a constraint language to describe symbolic states, in that case it will be a constraint language of the file system tree. The constraint language we need is still to be determined, in cooperation with the partners of the CoLiS project. The symbolic interpreter to design in this post-doc position is expected to be somehow generic, i.e. modular, with respect to the constraint language, in the same vein that formally verified abstract interpreters were recently designed [4] in the Verasco project (http://verasco.imag.fr/wiki/Main_Page). The symbolic execution approach will allow us to automatically infer for each script what are...
the preconditions on the file system (such as assumed existence of some files or directories) that are needed to guarantee execution without errors. It should also automatically compute post-conditions that are provided by the script execution of a package, that thus may be assumed by the scripts of packages that depend on the former package. The second goal of the job is to contribute to the set-up of an infrastructure for automatic execution of the symbolic interpreter on all Debian installation scripts, in an adequate order with respect to packages dependencies. This approach should be also used for proving properties about composition of scripts, such as showing that installation followed by de-installation is more or less the identity.


Main activities

- Design a symbolic execution engine for CoLiS programs
- Formalize this engine in Why3 and prove it correct
- Design an experimental platform for analysis of installation scripts
- Write reports and submit research articles

All these activities will be conducted in collaboration with other CoLiS project members. Regular working groups are organized at IRIF laboratory in Paris.

Skills

The candidate should have a PhD thesis in the domain of formal methods for software engineering. A basic knowledge of an interactive (such as Coq, Isabelle, PVS) or autoactive verification system (such as Why3, Dafny, KeY) is expected. Some taste in implementing using functional programming (OCaml, Haskell, etc.) would be a plus.

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

Remuneration

Monthly gross salary: 2,653 euros