# Title: Interaction Specification Mining

## Field:

Specification mining, AI, Formal Methods, Rewriting Techniques, Cybersecurity

**Expected duration:** – 4/6 months internship -

**Level:** Master engineering or science

## Description of the internship position

The appetite of the society for new technological features leads companies to create ever more complex systems composed many software-components in ever less time to market. This fast innovation tempo has often a negative impact in terms of software and system documentation. This makes such systems hard to validate and protect against vulnerabilities and cyberattacks, as there are no behavioral specifications that could typically serve as a reference oracle in a testing process or allow conducting security analyses of the possible attack.

An example of such systems are client/server systems, transportation control systems, Internet of Things, connected autonomous vehicles, etc. Therefore, discovering their formal specifications, i.e., specification mining [AmmonsBL02,BelluccinNT20], is of great interest for many software Verification and Validation (V&V) activities.

The objective of the internship is to **develop new algorithm for mining specifications from execution logs of communicating software components.** These logs are collected via code instrumentation or sniffing or via a testing architecture. Logged executions are typically sequences of actions occurring at interface of the components, which are expected to implement some communication protocol.

In this work, the mining will target specification, which are interactions such a UML Sequence Diagrams (UML-SD) or Message Sequence Charts (MSC). It will based on a recent work that provides expressive interactions (include rich scheduling and choice operators) with rewriting-based operational unfolding (defining transitions between interaction terms upon occurrence of send/receive action) and implemented into the tool HIBOU [MaheBGLL21, Mahe21].

The interaction mining process will apply rewriting strategies based on [Mahe21] to infer concise generalizing interactions from the execution logs of the communicating components. We will rely on the algebraic properties of the interactions, either in relation to the scheduling operators themselves (associativity, commutativity, neutral element), or to the projection mechanisms from a global system to its components.

**The specification mining algorithm will be developed in Rust the programming language used to implement the HIBOU tool** https://github.com/erwanM974/hibou_label.

## Location

The host unit in CEA will be the LECS team: development of methods and tools for the engineering of system requirements and compliance. CEA LIST, Palaiseau, Paris Area

## References


[AmmonsBL02] G. Ammons, R. Bodik, J. R. Larus. Mining specifications, POPL, 2002.

[BelluccinNT20] S. Belluccini, R. De Nicola, B. Re, and F. Tiezzi. PALM: A technique for process algebraic specificationmining, IFM 2020.

[MaheBGLL21] E. Mahe, B. Bannour, C. Gaston, A. Lapitre, P. Le Gall. A small-step approach to multi-trace checking against interactions, *SAC, ACM,* 2021.

[Mahe21] E. Mahe. An operational semantics of interactions for verifying partially observed executions of distributed systems. PhD thesis, University of Paris-Saclay, France, 2021.


**Candidate profile.** The candidate should have a strong background in Computer Science with excellent programming skills. Ideally, the candidate is equally interested by software development and solving research questions. Some knowledge of formal methods be it rewriting techniques, formal modeling (automata theory, process algebra, etc.), model learning and model-based testing will be appreciated.

## Specific Conditions.

Stipend will support internship daily life expenses, accomodation and local public transportation (limited to Paris area)

## Supervisor / Contacts.


Boutheina BANNOUR, PhD, Researcher at CEA LIST
boutheina.bannour@cea.fr

Pascale Le Gall, Professor at CentraleSupélec
pascale.legall@centralesupelec.fr