Proposition de stage

Deep Learning pour la sélection intelligente de stratégies d'analyse de code

Mots clés : Intelligence Artificielle, Machine learning, Interprétation Abstraite

Cadre du stage

Le CEA LIST est un centre de recherche technologique sur les systèmes à logiciel prépondérant qui mène ses recherches en partenariat avec les grands acteurs industriels (nucléaire, automobile, aéronautique, défense et médical) pour étudier et développer des solutions innovantes adaptées à leurs besoins. Au sein du CEA LIST, le Laboratoire Sûreté des Logiciels (LSL), localisé à Palaiseau (Essonne), développe les outils d'aide à la validation et à la vérification de logiciels et de systèmes matériels / logiciels, tout particulièrement dans les domaines des systèmes embarqués critiques et de la cybersécurité.

L'un des nos outils, nommé Frama-C (http://frama-c.com), est une plate-forme logicielle *open-source* développée en OCaml, facilitant le développement d'analyses de programmes C. Le stage se déroulera au sein de l'équipe développant Frama-C.

Objectifs du stage

Frama-C intègre de nombreuses techniques d'analyse statique, notamment dans l'un de ses plugins, Eva, dédié à la recherche de propriétés de programme et en particulier à la recherche d'erreurs à l'exécution (débordements arithmétiques, accès mémoire invalides et autres comportements indéfinis du C). Ces techniques sont efficacement combinées, mais leur activation peut coûter cher en termes de temps d'analyse. Il convient donc de choisir judicieusement quelles techniques utiliser dans quel cas.

Ces choix sont aujourd'hui réalisés par l'analyste, qui utilise l'outil afin de réaliser des preuves de sûreté sur un logiciel. Ils nécessitent une connaissance fine de Frama-C et de ses techniques d'analyse, tous deux évoluant avec le temps. En outre, pour être le plus fin possible, ces choix peuvent être réalisés à l'échelle de la fonction ou de la boucle dans un code source parfois grand. Cette double difficulté, reposant à la fois sur la quantité de connaissances et sur le temps nécessaire, peut s'avérer être un obstacle bloquant.

Pour lever cet obstacle, nous avons mis en place des techniques prometteuses de machine-learning. Il s'agit de traiter la difficulté non plus comme un problème de compréhension de l'outil d'analyse mais comme un problème de compréhension des formes de code adaptées aux différentes techniques. Ceci a deux avantages. D'une part, l'apprentissage peut être répété à mesure que l'outil évolue, afin de rester adapté aux caractéristiques des dernières techniques disponibles. D'autre part, le choix automatique des techniques peut être réalisé à un niveau beaucoup plus fin que ce qu'un humain peut accomplir en un temps raisonnable.

Les objectif du stage sont de contribuer à notre chaîne d'apprentissage spécialisée dans le code source, de participer à son intégration dans Frama-C et d'améliorer ses performances dans la sélection automatique de stratégies d'analyses :

- explorer les techniques d'apprentissage spécialisées sur les graphes,
- développer des représentations intermédiaires prenant en compte les flots de données et les flots de contrôle,
- explorer les techniques d'apprentissage en présence de données déséquilibrées et
- exploiter les prédictions obtenues dans Eva.

Candidatures

- **Profil**
  - Étudiant niveau master ou en deuxième ou troisième année d'école d'ingénieur
  - Connaissance d'au moins un langage impératif, de préférence le langage C
  - La connaissance du langage OCaml est un plus
  - Capacité de travail en équipe
- **Durée** : 5 à 6 mois
- **Conditions** : stage indemnisé, aide au logement possible, transports CEA en Île-de-France.
- **Encadrement** : Michele Alberti (*michele.alberti@cea.fr*) et Valentin Perrelle (*valentin.perrelle@cea.fr*)

Les délais administratifs de recrutement au CEA étant de 2 à 3 mois minimum, merci de prendre contact le plus tôt possible.