Rétro-ingénieurie : Décompilation de code assembleur

Frédéric Recoules and Sébastien Bardin

CEA, List

frederic.recoules@cea.fr sebastien.bardin@cea.fr
https://binsec.github.io/

La décompilation consiste à retrouver un code source (par exemple, du C d'assez haut niveau) à partir de langage assembleur ou même de code machine. L'outil BINSEC/Tina [1] développé au CEA List est un outil de décompilation pour assembleur "inline", c'est à dire de l'assembleur intégré dans du code C au moyen de primitives dédiées (supportées par GCC et Clang). L'outil a fait ses preuves pour ce type d'assembleur, et a par exemple permis de décompiler avec succès des milliers de codes asm inline pris dans des distributions Linux Debian.

Le but du stage sera d'une part d'étudier comment BINSEC/Tina peut être utilisé non plus sur du code assembleur inline mais sur des fonctions complètement écrites en assembleur, et d'autre part de remplacer le front-end C actuel de BINSEC/Tina par un frontend LLVM/Clang afin de gagner en robustesse.

Éléments logistiques. Le stage sera hébergé sur le site Nano-INNOV à Saclay. Le stage sera co-encadré par Frédéric Recoules et Sébastien Bardin.

Postuler. Pour candidater ou pour obtenir plus d'information, merci de prendre contact par email avec les encadrants.

Références

[1] Frédéric Recoules, Sébastien Bardin, Richard Bonichon, Laurent Mounier, and Marie-Laure Potet. Get rid of inline assembly through verification-oriented lifting. In *Proceedings of the 34th IEEE/ACM International Conference on Automated Software Engineering*, 2020.