

Adaptation en C/C++ d'un outil d'analyse statistique de bases de donnée de mots de passe

Encadrants : Patrick Lacharme et Mathieu Valois

Lorsque des hackers s'attaquent à des bases de donnée de mots de passe, il leur est très utile d'en faire ressortir les propriétés statistiques des mots trouvés : longueur minimum, maximum, jeux de caractères utilisés, etc. Chacun peut produire son propre outil de son côté, mais ceci est long et fastidieux. Il existe plusieurs projets dans différents langages, mais ces outils sont codés dans des langages interprétés (Python et Ruby), ce qui ralentit fortement l'analyse de grosses bases de plusieurs gigas octets voire dizaines de gigas octets.

Les objectifs principaux de ce projet sont :

- Étudier ces solutions en les comparant dans un court résumé des avantages et inconvénients de chacun et en les exécutant sur des bases de données de tailles différentes afin de comparer leurs performances
- Choisir le plus intéressant d'entre eux et réécrire ses fonctionnalités dans un langage compilé machine au choix : C, C++, Golang ou autre.

Si l'état d'avancement le permet, ou si décidé dès le début, rendre le code parallélisable. L'idéal est de concevoir dès le début la parallélisation du projet, puisque cela demande un grand remaniement du code et même parfois de la conception complète du programme.

Ressources :

- Pack (<https://thesprawl.org/projects/pack/>)
- Passpal (<http://thepasswordproject.com/passpal>)
- Pipal (<https://digi.ninja/projects/pipal.php>)