Sémantique et extension du langage C2QL pour la composition de techniques protégeant la confidentialité dans le nuage

Santiago Bautista

Juillet 2017

Résumé

Des applications de tout genre manipulent des personnelles de ses utilisateurs et utilisent le cloud pour s'exécuter ou s'héberger. Différentes techniques existent pour protéger la confidentialité de ces données. Pendant ce stage on a étudié la sémantique et prouvé les propriétés algébriques d'un langage permettant de décrire efficacement la composition de plusieurs de ces techniques, comme la fragmentation et le chiffrement.

Mots clés : privacy, cloud-computing, semantics, proof of correctness, algebraic laws, fragmentation, optimisation

Table des matières

1 Introduction

De plus en plus de logiciels sont développés pour être exécutés dans le cloud, et ses logiciels, de quelque sorte qu'ils soient (messagerie, gestion d'agenda personnel, commande de pizza ou reconnaissance vocale) traitent des données personnelles, qu'ils doivent donc protéger.

Une des propriétés qui doit être garantie dans la protection des données personnelles est la confidentialité. Différentes techniques existent pour protéger la confidentialité des données, comme par exemple le chiffrement et la fragmentation.

Dans sa thèse de 2016, Ronan Cherrueau montre que chacune de ces techniques a des avantages et des inconvénients, mais qu'*en composant les différentes techniques ensemble* on peut profiter de tous les avantages de ces techniques en éliminant la plupart des inconvénients. Il développe donc un langage, nommé C2QL (pour *Cryptographic Compositions for Query Language*), qui permet de décrire une telle composition de techniques de sécurisation des données pour en vérifier la correction et raisonner plus facilement.

Le langage se présente comme un ensemble de fonctions que l'on peut composer entre elles. Parmi ces fonctions, il y a les fonctions classiques pour faire des requêtes dans des bases de données, telles que la projection et la sélection, tout comme des fonctions décrivant la protection des données, comme le chiffrement ou la fragmentation.

Un des intérêts de ce langage est que, pour décider comment protéger les données des utilisateurs, le développeur peut suivre un processus simple en trois étapes. D'abord, le développeur écrit les requêtes *en ne tenant compte* ni du fait que le programme s'exécute dans le nuage, ni des mécanismes pour le protéger. Puis, il compose sa requête avec les fonctions de protection nécessaires (qui dépendent du problème en particulier, des contraintes de confidentialité spécifiques) pour avoir une requête sécurisée. Finalement, le développeur utilise des lois de commutation entre les différentes fonctions pour optimiser sa requête sécurisée.

Par conséquent, disposer de lois qui indiquent à quelles conditions les différentes fonctions du langage commutent est très important.

Pendant ce stage, j'ai complété l'ensemble de lois fournies dans la thèse de Ronan et j'ai formalisé la sémantique des différentes fonctions pour ensuite démontrer la correction de ces lois.