CRYPI – Projet 2023

10 / 05 / 2023

Théo Le Bever
Maxence Crouzy
Marius Andre
Romain Nakusi

Vue d'ensemble du calcul sécurisé multipartite :

Le calcul sécurisé multipartite (SMC) est une technique cryptographique qui permet à plusieurs parties de collaborer pour effectuer un calcul tout en préservant la confidentialité des données qu'ils manipulent. Dans ce projet, nous avons utilisé le SMC pour effectuer une régression linéaire en préservant la confidentialité des données et du modèle, en nous concentrant sur des paramètres semi-honnêtes. Un exemple classique de l'application du SMC est le problème des deux millionnaires qui souhaitent comparer leur richesse sans révéler le montant exact qu'ils possèdent.

Protocole SMC choisi et implémentation en Rust :

Le choix du langage Rust pour ce projet est dû à sa performance, sa sécurité et sa gestion optimisée de la mémoire était quelque chose d'essentiel pour nous. De plus, le projet utilise le cadre gRPC de Tonic pour permettre une communication sécurisée entre le propriétaire des données et le propriétaire du modèle. En effet, cela nous permettait une intégration simple et efficace du protocole TLS.

Afin de générer des certificats et ainsi établir la communication sécurisée entre le client et le serveur, un script shell est utilisé. De plus, il permet également la génération de clés RSA de 2048 bits qui sont-elles utilisées pour assurer une sécurité robuste. La taille de la clé est un élément crucial pour garantir la résistance aux attaques cryptanalytiques, et une clé de 2048 bits est actuellement considérée comme offrant un niveau de sécurité adéquat. Ces clés sont ensuite utilisées pour créer les certificats et établir une connexion TLS sécurisée entre les parties, assurant ainsi la confidentialité et l'intégrité des données échangées. Le propriétaire du modèle partage le modèle de régression linéaire, tandis que le propriétaire des données fournit les données d'entrée pour la prédiction. La prédiction et l'évaluation du modèle sont effectuées sans révéler d'informations sensibles à l'une ou l'autre des parties.

Sélection de la bibliothèque HE/SMC :

Pour garantir la compatibilité avec le langage Rust et assurer une communication sécurisée, la bibliothèque Tonic a été choisie en raison de sa prise en charge de gRPC et de TLS. Le projet utilise également le protocole HMAC pour le calcul de hash et SHA-256 pour la vérification d'intégrité, garantissant ainsi la sécurité des données échangées.

Une CI est également en place sur Github afin d'effectuer une vérification des bibliothèques sélectionnées de manière automatique à l'aide de l'utilitaire cargo.

Résultats des expériences :

Les expériences menées ont montré que des prédictions précises peuvent être obtenues à partir du modèle de régression linéaire en utilisant le protocole SMC, tout en préservant la confidentialité des données et du modèle. La précision du modèle de prédiction sécurisé était comparable à celle du modèle de prédiction non sécurisé.

Défis rencontrés :

Parmi les défis rencontrés au cours du projet, on peut citer l'apprentissage du langage Rust tout d'abord. Tous les membres de l'équipe ne connaissaient pas encore ce langage de programmation et ont dû l'apprendre rapidement. De plus, la communication réseau a évidemment été un point d'attention puisque cela représente tout le cœur du projet. Enfin, le choix des protocoles utilisées dans le projet a été une considération importante puisque ce projet se devait d'avoir un équilibre entre performance et sécurité.

Recommandations pour l'avenir :

Un axe d'amélioration de ce projet serait l'implémentation des modèles et des algorithmes d'apprentissage automatique plus complexes. De plus, aujourd'hui le projet protège les utilisateurs des attaques extérieures. Cependant, le projet ne prend pas en charge un environnement totalement malveillant, dans lequel les parties peuvent s'écarter du protocole. Ce serait donc un point d'amélioration intéressant.

Conclusion :

Le calcul sécurisé multipartite permet de réaliser des opérations sur des données tout en préservant leur confidentialité. Il a de nombreuses applications, notamment dans l'utilisation de données à caractères personnels tel que les données médicales. Ce projet démontre que le SMC peut être appliqué avec succès lors de cas simple en utilisant par exemple la régression linéaire pour préserver la confidentialité des données et du modèle tout en maintenant une précision de prédiction élevée. Le choix de Rust et de la bibliothèque Tonic a permis d'assurer une communication sécurisée entre les parties et une implémentation optimisée du protocole SMC.

En somme, cette étude met en évidence l'importance et la pertinence du calcul sécurisé multipartite dans le domaine de l'apprentissage automatique et de la protection de la vie privée. Les travaux futurs dans ce domaine devraient se concentrer sur l'amélioration des protocoles existants, l'exploration de nouvelles bibliothèques et méthodes, et l'adaptation des solutions SMC à un éventail plus large de problèmes d'apprentissage automatique. En fin de compte, ces efforts contribueront à créer des systèmes de traitement de données plus sécurisés et plus respectueux de la vie privée, ce qui est essentiel dans notre monde de plus en plus connecté et dépendant des données.