SIESTA - Cloud confidentiel pour la recherche

Vincent Legoll*[1] and Jérôme Pansanel[2]

[1]Institut Pluridisciplinaire Hubert Curien – CNRS : UMR7178, Universié de Strasbourg – France
[2]Institut Pluridisciplinaire Hubert Curien (IPHC) – CNRS : UMR7178, Universié de Strasbourg – 23, rue du Loess - BP28 - 67037 STRASBOURG CEDEX 2, France

Résumé

Les chercheurs des différentes disciplines scientifiques traitent de nombreux types de données, dont certaines sont considérées comme sensibles. En effet, elles peuvent contenir des données personnelles, des secrets industriels ou brevetables, etc.
La législation, tant à l'échelle nationale, qu'européenne ou internationale protège ces données personnelles et les lois ne permettent plus leur exploitation incontrôlée.
Il est donc nécessaire de mettre à disposition des outils pour les chercheurs afin d'analyser et de traiter ce type de données.

Les données peuvent être sécurisées lors de leur stockage ou des transferts, par des moyens cryptographiques, mais jusqu'à présent, lors de leur utilisation les données sont disponibles en mémoire sous forme non cryptée. Les administrateurs de tels systèmes peuvent donc accéder à toutes les données présentes lors de leur traitement, en accédant directement au contenu de la mémoire de la machine.

Les systèmes d'informatique en nuage (cloud), bien qu'intéressants sur le plan de la souplesse, ajoutent une incertitude quant à la possibilité pour des personnes non autorisées d'accéder à ces données sensibles.

Le secteur privé a déjà compris l'importance de cette problématique et une offre existe chez tous les grands fournisseurs de ressources cloud.

Le projet européen SIESTA vise à mettre en place des systèmes de cloud confidentiels, à savoir, des systèmes avec lesquels les administrateurs d'une plateforme matérielle n'ont pas de moyen d'accéder aux données en cours d'utilisation, car situées dans des enclaves sécurisées dont tout l'environnement virtuel est crypté et supervisé.

Les constructeurs de matériel (CPU) ont développé des fonctionnalités qui permettent de mettre en place le cryptage des données en mémoire vive, ainsi que le contexte d'exécution du processeur. Les solutions matérielles existent chez AMD, ARM, Intel, a divers degrés de fonctionnalités ou d'intégration dans les piles logicielles FOSS.

Le projet technique comporte plusieurs sujets :

*Intervenant

* authentification et autorisation (utilisateurs, rôles, droits d'accès)
les enclaves sécurisées (conteneurs ou machines virtuelles)
le stockage sécurisé
le réseau
le développement
l'audit de sécurité (vulnérabilités, logs, surveillance)

Pour aider à cerner les besoins techniques, le projet SIESTA s'est adjoint plusieurs cas d'usage dans des domaines scientifiques variés :

* Épidémiologie
Imagerie médicale
Énergie
Anonymisation de données textuelles automatisée
Démographie

Le projet SIESTA reste lié aux principes FAIR, même en cas de données confidentielles, et les problématiques liées sont prises en compte par la plateforme.

Il existe d'autres initiatives sur des sujets similaires ou connexes :
* EOSC ENTRUST
EUCAIM
TITAN

Le projet étant en cours, cet exposé permettra de présenter la première version de la plateforme, ainsi que les évolutions prévues pour la seconde phase du projet, qui permettra de mettre à disposition des chercheurs un service de niveau production (TRL 8). Pour cela, un suivi des utilisateurs est mis en place afin de déterminer les sujets qui méritent d'être approfondis ou améliorés.