Offre de stage M2 - Vers la mise en oeuvre de systèmes d'apprentissage distribués collaboratifs et sécurisés

Mots-clés : Environnements d'Exécution de Confiance, Intelligence Artificielle, Apprentissage Fédéré, Sécurité, Rémunération

Organisme d'accueil : Laboratoire COSYS-ERENA, Université Gustave Eiffel, Bordeaux, France

Encadrants :

- Yannis Formery (yannis.formery@univ-eiffel.fr)
- Léo Mendiboure (leo.mendiboure@univ-eiffel.fr)

Profil attendu :

- Niveau Bac + 5 en Informatique/Télécommunications/Traitement du Signal/Cybersécurité (École d'ingénieur ou Master)
- Connaissance de base en Réseaux de Communication et Architectures cellulaires
- Connaissance de base en Intelligence Artificielle
- Autonomie et capacité d'adaptation à un environnement orienté vers la recherche
- Maîtrise de l'anglais (oral et écrit)

Informations complémentaires :

- Durée : 5 à 6 mois
- Démarrage : À partir de mi-février ou mars 2024

Contexte

L'application d'outils de l'Intelligence Artificielle (IA) touche aujourd'hui absolument tous les domaines, qu'il s'agisse de santé, de systèmes de transport, d'agriculture ou encore d'éducation [1]. Parmi les solutions émergentes, l'Apprentissage Fédéré (Federated Learning) figure aujourd'hui parmi les approches les plus en vogue [2]. Son principe est simple : les données restent localisées sur les appareils des utilisateurs (par exemple, le serveur d'un hôpital, le siège d'une entreprise, une station de base ou un téléphone), et seuls les poids des modèles mis à jour sont partagés avec d'autres. Cela permet d'entraîner des modèles globaux (combinant les informations provenant de différents centres, utilisateurs, etc.) sans centraliser les données, renforçant ainsi la protection de la vie privée. Dans cette architecture, certains composants jouent un rôle central : les agrégateurs. Ces

derniers reçoivent les poids des modèles des utilisateurs ou clients, puis leur renvoient un modèle global résultant d'une pondération entre leurs contributions respectives.

Ce projet considère donc le cas suivant : un système décentralisé dans lequel toute personne peut proposer ses ressources pour faire office d'agrégateur de modèles de clients. Cela permet à ces personnes de tirer parti des capacités calculatoires dont elles disposent, puisqu'elles sont rémunérées pour les opérations d'agrégation qu'elles réalisent. Ce modèle dispense également les clients du besoin de déployer des infrastructures spécifiquement dédiées à l'agrégation de modèles, ce qui peut potentiellement entraîner une réduction des coûts.

Dans ce contexte, la question centrale est la suivante : comment mettre en œuvre ce type de système tout en garantissant la sécurité des échanges et le bon fonctionnement du service ? Cela inclut l'ensemble des étapes, depuis la remontée des modèles, le choix des agrégateurs, l'agrégation elle-même, jusqu'à la vérification du travail des agrégateurs et leur rémunération.

Pour répondre à ces défis, ce projet propose de s'appuyer sur des technologies matérielles de sécurité, telles que les HSM (Hardware Security Modules) et les environnements d'exécution de confiance TEE (Trusted Execution Environments) [3]. Ces solutions permettent d'assurer la confidentialité et l'intégrité des modèles, ainsi que la sécurisation des calculs et des échanges. Les HSM fournissent un stockage sécurisé des clés cryptographiques et une gestion des opérations sensibles, tandis que les TEE garantissent que les calculs sont exécutés dans un environnement isolé et protégé contre les attaques extérieures [4].

Missions

L'objectif de ce stage sera donc de concevoir et de développer un **Proof-of-Concept (PoC)** démontrant le potentiel d'un tel système d'IA décentralisé, sécurisé par l'utilisation des HSM et des TEE. Ce système repose sur une communauté, où chaque participant peut jouer un rôle actif tout en bénéficiant d'une protection avancée de ses données et de ses modèles.

Pour parvenir à cet objectif différentes étapes seront nécessaires :

- Appréhender la documentation liée au contexte considéré (HSM/TEE par exemple) ;
- Mettre en place le système décrit dans la section I : sous la supervision de vos encadrants, vous devrez sélectionner les composants nécessaires pour le déploiement de la solution d'apprentissage fédéré ainsi que pour la mise en œuvre des technologies de sécurisation basées sur les HSM/TEE. Vous serez chargé(e) d'interconnecter les différents éléments identifiés, notamment les frameworks d'IA et les mécanismes de sécurité, pour assurer une intégration cohérente et fonctionnelle.
- Développer et mettre en œuvre les mécanismes fondamentaux du système : Sélection des agrégateurs, Transfert sécurisé des données entre clients et agrégateurs, Agrégation des modèles, Transmission des résultats, Validation et rémunération des agrégateurs. Ces tâches devront être réalisées en prenant en compte les contraintes spécifiques des technologies HSM et TEE, telles que les limitations en termes de puissance de calcul, de bande passante ou de stockage sécurisé

- Définir, avec vos encadrants, un cas d'usage simple mais représentatif pour démontrer le fonctionnement de la solution développée. Vous serez également responsable de mettre en place une démonstration finale, qui pourra servir de preuve de concept (PoC) pour évaluer la faisabilité et l'efficacité du système proposé.
- Rédiger un rapport présentant les principales conclusions de cette étude et une potentielle valorisation au travers de la rédaction d'un article de recherche.

Références

[1] Rao, V. S., Satish, M. A., & Prasad, M. B. (2024). *Artificial intelligence: Principles and applications*. Leilani Katie Publication.

[2] Bonawitz, K. (2019). Towards federated learning at scale: System design. *arXiv preprint arXiv:1902.01046*.

[3] Sabt, M., Achemlal, M., & Bouabdallah, A. (2015, August). Trusted execution environment: What it is, and what it is not. In *2015 IEEE Trustcom/BigDataSE/Ispa* (Vol. 1, pp. 57-64). IEEE.

[4] Li, J., Chen, N., Yu, S., & Srivatanakul, T. (2024, October). Efficient and Privacy-Preserving Integrity Verification for Federated Learning with TEEs. In *MILCOM 2024-2024 IEEE Military Communications Conference (MILCOM)* (pp. 999-1004). IEEE.