الجمهورية الشعبية الديمقراطية الجزائرية
democratic and popular republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي ٠80 ماي 1945٠ بسيدي بلعباس
Higher School of Computer Science 08 May 1945 - Sidi Bel Abbes

Thesis of End of Study

In view of obtaining the diploma of **state engineer**
Field: **Computer science**

Specialty: **Computer systems engineering**

Enhancing Federated Learning Research through NS3-FLSim and Model Poisoning Mitigation

---

Presented by:

- Abdelilah KECHIDI
- Ayoub KADDOUR

Presented on: ../09/2023

In front of the jury composed of:

- **M. …….** President
- **M.AZZA Mohammed** Supervisor
- **M. …….** Examinator

Academic Year: 2022/2023

Abstract

In recent years, the dynamic landscape of machine learning has been significantly reshaped by the rapid emergence of federated learning, marking a transformative paradigm shift that addresses critical challenges inherent in traditional centralized machine learning approaches. This groundbreaking methodology empowers multiple decentralized devices or entities to collaboratively enhance a shared model, all while avoiding the need to directly exchange raw local data. This unique characteristic of federated learning not only champions data privacy but also confronts issues surrounding distributed data utilization and the cumbersome communication overhead that often plagues centralized models.

The allure of federated learning lies in its capacity to harmonize the efficient utilization of data spread across a myriad of devices or entities, all while meticulously upholding the tenets of privacy and data ownership. This revolutionary approach acknowledges the sensitivities of data ownership and privacy concerns that have grown increasingly pronounced in our digitally interconnected world. By enabling diverse entities to partake in collaborative model enhancement without compromising sensitive data, federated learning navigates the intricate path between the demand for data-driven insights and the imperative to preserve individual data sovereignty.

In the contemporary landscape, where data is often likened to a precious commodity, federated learning has swiftly ascended to a position of paramount importance. It constitutes an unparalleled solution, offering not only privacy preservation on an unprecedented scale but also an avenue for forging collaborative ties among dispersed entities. The very essence of federated learning lies in its ability to tailor models to the idiosyncrasies of individual devices, effectively personalizing algorithms to device-specific attributes, thus bolstering the performance and efficacy of these models across a heterogeneous network of participants.

This dissertation meticulously probes the realm of federated learning within the specific context of NS3-FL, a pioneering federated learning simulator that takes into account the intricate interplay of data, algorithms, and network dynamics. By em-

barking on an investigative journey, this study delves into the aggregation methods currently employed within the NS3-FL framework, shedding light on their inherent limitations and underscoring the need for innovation. Through rigorous research and development, this dissertation pioneers novel aggregation techniques that seamlessly integrate into the NS3-FL ecosystem, breathing new life into the very core of federated learning.

However, the advancement of federated learning is not solely reliant on improved aggregation methods; it must also grapple with the escalating concern of security vulnerabilities and potential 5adversarial attacks. Recognizing this imperative, this dissertation introduces a pioneering security framework designed to fortify federated learning models against a gamut of potential attacks. By innovatively blending robust security mechanisms with the intricate fabric of federated learning, this study bolsters the foundation upon which the entire field is built, ensuring its viability and relevance in a world fraught with ever-evolving cybersecurity challenges.

In essence, this dissertation stands as a pivotal milestone in the journey toward realizing the true potential of federated learning. Through meticulous exploration, innovative aggregation methodologies, and a pioneering approach to security, this study emerges as a beacon guiding the field toward a future where collaborative, privacy-conscious, and secure machine learning is not just a possibility, but an actively flourishing reality.

Résumé

Ces dernières années, le paysage dynamique de l'apprentissage automatique a été considérablement remodelé par l'émergence rapide de l'apprentissage fédéré, marquant un changement de paradigme transformateur qui aborde les défis critiques inhérents aux approches traditionnelles de l'apprentissage automatique centralisé. Cette méthodologie révolutionnaire permet à de multiples dispositifs ou entités décentralisés de collaborer pour améliorer un modèle partagé, tout en évitant le besoin d'échanger directement des données locales brutes. Cette caractéristique unique de l'apprentissage fédéré défend non seulement la protection des données, mais confronte également les problèmes liés à l'utilisation des données distribuées et à la surcharge de communication fastidieuse qui affecte souvent les modèles centralisés.

L'attrait de l'apprentissage fédéré réside dans sa capacité à harmoniser l'utilisation efficace des données réparties sur une multitude de dispositifs ou entités, tout en respectant méticuleusement les principes de la vie privée et de la propriété des données. Cette approche révolutionnaire reconnaît les sensibilités de la propriété des données et les préoccupations en matière de confidentialité qui se sont de plus en plus prononcées dans notre monde numériquement interconnecté. En permettant à des entités diverses de participer à l'amélioration collaborative d'un modèle sans compromettre les données sensibles, l'apprentissage fédéré parcourt le chemin complexe entre la demande d'informations basées sur les données et l'impératif de préserver la souveraineté des données individuelles.

Dans le paysage contemporain, où les données sont souvent assimilées à une précieuse matière première, l'apprentissage fédéré a rapidement atteint une position d'importance primordiale. Il constitue une solution inégalée, offrant non seulement la préservation de la vie privée à une échelle sans précédent, mais également une voie pour forger des liens collaboratifs entre des entités dispersées. L'essence même de l'apprentissage fédéré réside dans sa capacité à adapter les modèles aux spécificités des dispositifs individuels, en personnalisant efficacement les algorithmes en fonction des attributs spécifiques des dispositifs, renforçant ainsi la performance et l'efficacité de ces modèles au sein d'un réseau hétérogène de participants.

Cette thèse explore méticuleusement le domaine de l'apprentissage fédéré dans le contexte spécifique de NS3-FL, un simulateur pionnier d'apprentissage fédéré qui prend en compte l'interaction complexe entre les données, les algorithmes et la dynamique du réseau. En entreprenant un voyage d'investigation, cette étude se penche sur les méthodes d'agrégation actuellement utilisées dans le cadre de NS3-FL, mettant en lumière leurs limites inhérentes et soulignant la nécessité d'innovation. Grâce à une recherche et un développement rigoureux, cette thèse propose des techniques d'agrégation novatrices qui s'intègrent parfaitement à l'écosystème NS3-FL, insufflant ainsi une nouvelle vie au cœur même de l'apprentissage fédéré.

Cependant, l'avancement de l'apprentissage fédéré ne repose pas uniquement sur des méthodes d'agrégation améliorées ; il doit également faire face à la préoccupation croissante des vulnérabilités en matière de sécurité et des attaques adverses potentielles. Consciente de cette nécessité, cette thèse introduit un cadre de sécurité pionnier conçu pour renforcer les modèles d'apprentissage fédéré contre toute une gamme d'attaques potentielles. En mélangeant de manière innovante des mécanismes de sécurité robustes avec la structure complexe de l'apprentissage fédéré, cette étude renforce les bases sur lesquelles repose l'ensemble du domaine, assurant sa viabilité et sa pertinence dans un monde riche en défis en constante évolution en matière de cybersécurité.

En essence, cette thèse constitue une étape cruciale dans le voyage vers la réalisation du véritable potentiel de l'apprentissage fédéré. Grâce à une exploration méticuleuse, à des méthodologies d'agrégation innovantes et à une approche pionnière en matière de sécurité, cette étude se présente comme un phare guidant le domaine vers un avenir où l'apprentissage automatique collaboratif, respectueux de la vie privée et sécurisé, n'est pas seulement une possibilité, mais une réalité en plein essor.

مُلخص

في السنوات الأخيرة، تم إعادة تشكيل البيئة الديناميكية للتعلم اآللي بشكل كبير بفضل ظهور سريع للتعلم التفاعلي، مما يشكل تحوالًنمطيًا يعالج التحديات الحرجة المترتبة على النهج التقليدي للتعلم اآللي المركزي. تمنح هذه المنهجية الرائدة العديد من األلجهزةأو الكيانات غير المركزية القدرة على تعزيز نموذج مشترك بالتعاون، دون الحاجة إلى تبادل البيانات المحلية الخام مباشرة.هذه الخاصية الفريدة في التعلم التفاعلي ال تحمي فقط خصوصية البيانات، ولكنها تواجه أيضًا مشاكل تتعلق باالستفادة منالبيانات الموزعة والعبء الزائد لالتصال الذي يعاني منه غالبًا النماذج المركزية.

جاذبيةالتعلم التفاعلي تكمن في قدرته على تنسيق االستفادة الفعالة من البيانات المنتشرة عبر مجموعة متنوعة من األلجهزة أو الكيانات،مع الحفاظ بعناية على مبادئ الخصوصية وملكية البيانات. تعترف هذه النهج الثوري بحساسيات ملكية البيانات والمخاوفالمتزايدة بشكل ملح في عالمنا المتصل رقميًا. من خالل تمكين الكيانات المتنوعة من المشاركة في تعزيز النموذج بالتعاوندون المساس بالبيانات الحساسة، يستكشف التعلم التفاعلي المسار المعقد بين الطلب على التحليالت القائمة على البيانات وضرورةالحفاظ على سيادة البيانات الفردية.

فيالمناظرة المعاصرة، حيث يُشبه البيانات في كثير من األلحيان مادة قيمة، ارتقى التعلم التفاعلي بسرعة إلى موقع أهمية قصوى.إنه يشكل حالً ال مثيل له، حيث يقدم ليس فقط الحفاظ على الخصوصية على نطاق غير مسبوق، ولكن أيضًا وسيلة لبناء عالقاتتعاونية بين الكيانات المتناثرة. جوهر التعلم التفاعلي يكمن في قدرته على تخصيص النماذج لتفاصيل أجهزة الفردية، ممايزيد من أداء وفعالية هذه النماذج عبر شبكة متنوعة من المشاركين.

تستقصي هذه الرسالة بدقة مجال التعلم التفاعلي في سياق محدد وهو NS3-FL ،وهو محاكي رائد للتعلم التفاعلي يأخذ في اعتبارهالتفاعل المعقد بين البيانات والخوارزميات وديناميات الشبكة. من خالل االنطالق في رحلة تحقيقية، تتناول هذه الدراسة أساليباجمع المستخدمة حاليًا ضمن إطار NS3-FL ،مسلطة الضوء على قيودها الكامنة وتسليط الضوء على الحاجة إلى االبتكار.من خالل البحث والتطوير الدقيق، تقدم هذه الرسالة تقنيات جمع جديدة تتكامل بسالسة في نظام NS3-FL ،مما يعيد الحياةإلى نواة التعلم التفاعلي.

ومع ذلك، ليس تقدم التعلم التفاعلي معتمدًا على أساليب الجمع المحسنة فقط؛ بل يجب أيضًا التعامل مع المخاوف المتصاعدة بشأنثغرات األلمان والهجمات العدائية المحتملة. من خالل االعتراف بهذا األلم الضروري، تقدم هذه الرسالة إطارًا أمنيًا رائدًا مصمً مالتعزيز نماذج التعلم التفاعلي ضد مجموعة متنوعة من الهجمات المحتملة. من خالل دمج آليات األلمان القوية بشكل مبتكرمع نسيج التعلم التفاعلي المعقد، تعزز هذه الدراسة االسس التي يتم بناء الحقل بأكمله عليها، مضمنة استدامته واستمراريتهفي عالم مليء بتحديات أمن المعلومات المتطورة باستمرار.

باألساس،تقف هذه الرسالة كمعلم بارز في رحلة تحقيق اإلمكانيات الحقيقية للتعلم التفاعلي. من خالل استكشاف دقيق ومنهجي ألساليباجمع المبتكرة ومنهجية رائدة في مجال األلمان، تظهر هذه الدراسة كمشعل يرشد الميدان نحو مستقبل حيث يكون التعلماآللي التعاوني والمحترم للخصوصية واآلمن ليس مجرد احتمال، بل هو واقع مزدهر بالفعل.