Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science

Cybersecurity Enhancement of Federated Learning System Based on Confidential Consortium Framework(CCF)

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Doctor of Philosophy in
Computer Science**

By

Ibrahim Mohammed Ahmed Al halema

Supervised by

Prof. Dr. Manar Younis Kashmolaa

2023 A.D. **1444 A.H.**

ABSTRACT

Nowadays, intelligent systems attract a lot of attention for their services in all fields. Federated learning systems play an essential role in intelligent systems that operate in cyberspace. It is considered as one of the most prominent applications that provide intelligent services for the 6G. However, these intelligent systems face many issues in terms of security and reliability. The main issues that face federated learning are security and reliability, which could be affected by the poisoning attacks. In addition, poisoning attacks targeted models and data of federated learning systems. The most common poisoning attacks are either insider or outsider attacks. The outsider attacks are represented by impersonation attacks such as the Sybil attack. In contrast, insider attacks are represented by GAN attacks. So, providing a secure environment, which prevents insider and outsider attacks, considers this study's main aim.

This thesis proposed a new federated learning framework with a high-security level against Sybil and GAN poisoning attacks. The proposed framework, called FEDerated_CCF(FED_CCF), creates a hybrid environment using a federated learning system with Microsoft Confidential Consortium Framework (CCF); this can be achieved by constructing four containment zones. Each containment zone has a specific task that depends on multiple layers of security. The hybridity processes between federated learning and CCF has been implanted in the first containment zone. The first containment zone verifies participant devices' provenance by matching the device manifest with the FED_CCF network certificate. The joined devices gained trust after meeting the FED_CCF network requirements. BFT protocols and the Merkle tree control these devices throughout their working life in the FED_CCF system. The second containment zone is responsible for preparing the devices participating(worker devices) in the training operations by giving these trusted devices the local training model . in addition to the parameters of the last training round and the training dataset of the device. Moreover, the roles of these devices are assigned. Also, the third containment zone is responsible for validating the received parameters of all associated worker devices. The validator's device does this validation by examining the received device's

transaction using the Merkle tree and evaluating its parameters by calculating classification accuracy. In addition, examine the device network certificate and set the device status with a benign label for the trusted device or a malicious label for the untrusted device. Finally, the fourth containment zone miner's devices are responsible for aggregating associated worker device parameters with their validation reports. The miners in this containment zone select the best-associated worker based on the voting of the miner's devices.

The evaluation of the proposed FED_CCF is achieved by employing the MNIST dataset and analyzing the experimental results in two phases. The first phase describes the diagnostic accuracy of the used devices, either benign or malicious. The second phase represents the classification accuracy of the used dataset. These two experiments are applied to assess the performance of the suggested FED_CCF system for both insider and outsider attacks. The experiment results of the proposed FED_CCF system show 91.92% diagnosis accuracy and 96.26% classification accuracy against outsider attacks represented by Sybil attacks. Also, The experiment results of the proposed FED_CCF system show 90.73% diagnosis accuracy and 95.83% classification accuracy against insider attacks represented by GAN attacks. For more investigation, T-test has been used to show the FED CCF is the best in terms of mean classification and diagnosis accuracy in both Sybil and GAN attacks. From all the obtained results, it is clear that the performance of FED_CCF is superior compared with other methods from the literature and provides a secure environment for a federated learning system.

وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

تعزيز الأمن السيبراني لنظام التعلم الفيدرالي على أساس إطار عمل الاتحاد السري (CCF)

اطروحة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة دكتوراه فلسفة في
علوم الحاسوب

من قبل

ابراهيم محمد احمد الحليمة

بإشراف
أ.د. منار يونس كشمولة

١٤٤٤هـ ٢٠٢٣ م

المستخلص

في الوقت الحاضر ، تجذب الأنظمة الذكية الكثير من الاهتمام لخدماتها في جميع المجالات ، مع نمو العديد من التطبيقات الذكية. إذ تؤدي أنظمة التعلم الاتحادي دورًا أساسيًا في الأنظمة الذكية التي تعمل في الفضاء السيبراني ، و تعد من أبرز التطبيقات التي تقدم خدمات ذكية لجيل الشبكات السادس. لكن هذه الأنظمة تواجه العديد من المشكلات من حيث الأمان والوثوقية.

تواجه أنظمة التعلم الاتحادي تحديات كبيرة في مجالي الأمان والوثوقية عن طريق تأثرها بهجمات التسمم. تستهدفت هجمات التسمم النماذج والبيانات الخاصة بأنظمة التعلم الاتحادي. إن هجمات التسمم الأكثر شيوعًا هي إما هجمات من داخل النظام أو هجمات من خارج النظام. تعمل الهجمات الخارجية باسلوب هجمات انتحال الهوية مثل هجوم Sybil. في المقابل ، تعمل الهجمات الداخلية على تسميم البيانات من قبل الأجهزة الموثوقة و بواسطة هجمات GAN.

لذلك ، فإن توفير بيئة آمنة تمنع الهجمات الداخلية والخارجية يعد الهدف الرئيس لهذه الدراسة. لهذا السبب اقترحت هذه الأطروحة إطارًا جديدًا لانظمة التعلم الاتحادي يوفر مستوى أمان عالٍ ضد هجمات التسمم Sybil وGAN .

إطار العمل المقترح في هذه الدراسة هو FEDerated_CCF (FED_CCF) ، يمثل بيئة هجينة لتمثيل نظام التعلم الاتحادي الآمن ويتمتع بوثوقية الأجهزة المشاركة فيه عن طريق ربط نظام التعلم الاتحادي مع إطار الاتحاد السري Microsoft CCF ؛ من خلال إنشاء أربع مناطق احتواء. لكل منطقة مهمة محددة تعتمد على طبقات أمنية متعددة. تبداعمليات التهجين بين أنظمة التعلم الاتحادي و CCF في منطقة الاحتواء الأولى. حيث تتحقق المنطقة الأولى من مصدر الأجهزة المشاركة من خلال مطابقة بيان الجهاز بشهادة شبكة FED_CCF والأجهزة التي اكتسبت الثقة بعد تلبية متطلبات شبكة FED_CCF. يسيطر عليها بواسطة بروتوكولات BFT وشجرة Merkle طوال فترة عملها في النظام.

منطقة الاحتواء الثانية هي المسؤولة عن تجهيز الأجهزة المشاركة (أجهزة العاملين) في عمليات التدريب من خلال إعطاء هذه الأجهزة الموثوقة نموذج التدريب المحلية. فضلاً عن معلمات الجولة التدريبية الأخيرة ومجموعة بيانات التدريب الخاصة بالجهاز.علاوة على ذلك ، يتم تعيين أدوار هذه الأجهزة. أيضًا . منطقة الاحتواء الثالثة مسؤولة عن التحقق من صحة المعلمات المستلمة لجميع أجهزة العمال المرتبطة بها . يقوم جهاز المدقق بإجراء هذا التحقق من الصحة من خلال فحص معاملة الجهاز المستلم باستخدام شجرة Merkle وتقييم معلماته من خلال حساب دقة التصنيف. فضلاً عن ذلك ، يتم فحص شهادة شبكة الجهاز وتعيين حالة الجهاز مع تسميته اما جهاز حميد للأجهزة الموثوقة أو جهاز ضار للجهاز غير الموثوق به. أخيرًا ، تكون أجهزة عمال المناجم في منطقة الاحتواء الرابعة مسؤولة عن تجميع معلمات جهاز العامل المرتبطة بتقارير التحقق من الصحة الخاصة بهم. تختار المنطقة أفضل عامل مرتبط بناءً على تصويت أجهزة التعَدِّين. يمكن تحقيق تقييم FED_CCF المقترح بناءً على مجموعة بيانات MNIST من خلال مناقشة النتائج التجريبية في جزأين. يصف الجزء الأول دقة

التشخيص للأجهزة المستخدمة سواء كانت حميدة أو ضارة. يمثل الجزء الثاني دقة التصنيف لمجموعة البيانات المستخدمة. تم تنفيذ هاتين التجربتين لتقييم أداء نظام FED_CCF المقترح ضد الهجمات الداخلية والخارجية.

أظهرت نتائج التجربة لنظام FED_CCF المقترح دقة تشخيص بنسبة ٩١,٩٢٪ ودقة تصنيف بنسبة ٩٦,٢٦٪ ضد الهجمات الخارجية التي تمثلها هجمات Sybil. أيضًا ، تُظهر نتائج التجربة لنظام FED_CCF المقترح دقة تشخيص بنسبة ٩٠,٧٣٪ ودقة تصنيف ٩٥,٨٣٪ ضد الهجمات الداخلية التي تمثلها هجماتGAN. لمزيد من التحقيقات ، تم استخدام اختبار T لإظهار أن FED CCF هو الأفضل من حيث متوسط التصنيف ودقة التشخيص في كل من هجمات Sybil و GAN. من جميع النتائج التي تم الحصول عليها ، من الواضح أن أداء FED_CCF متفوق مقارنة بالطرق الأخرى من الأدبيات ويوفر بيئة آمنة لنظام التعلم الاتحادي.