Financiado por la Unión Europea
NextGenerationEU

MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

Plan de Recuperación,
Transformación
y Resiliencia

6G-RIEMANN-ML Entregable E12

Definition of the usecases for privacy preserving machine learning

PROGRAMA DE UNIVERSALIZACIÓN DE INFRAESTRUCTURAS DIGITALES PARA LA COHESIÓN UNICO I+D 5G 2021

Fecha: 31/7/2022

Versión: 1.0

Propiedades del documento

Id del documento	E12			
Título	Definition of the usecases for privacy preserving machine learning			
Responsable	UC3M			
Editor	Albert Banchs			
Equipo editorial	**Partner**	**Name**	**Surname**	**Sections**
	UC3M	Marco	Gramaglia	All
Nivel de diseminación	Público			
Estado del documento	Final			
Versión	1.0			

Historial

Revisión	Fecha	Por	Descripción
1.0	31/07/22	Editor	Final version

Revisor

Equipo revisor	**Partner**	**Name**	**Surname**	**Sections**
	UC3M	Albert	Banchs	All

Descargo de responsabilidad

This document has been produced in the context of the 6G-RIEMANN Project. The research leading to these results has received funding from the Spanish Ministry of Economic Affairs and Digital Transformation and the European Union-NextGenerationEU through the UNICO 5G I+D programme. The information contained in this document is provided "as is" without any express or implied warranties, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. The document writer shall not be liable for any damages, whether direct or indirect, arising out of or in connection with the use of this information. The user of this document assumes all risks and liabilities associated with its use and shall indemnify and hold harmless the document writer from any and all claims, losses, damages, or expenses, including attorney's fees, arising from the use of this information.

Table of Contents

Lista de acrónimos

MLaaS - Machine Learning as a Service

IoT - Internet of Things

AI - Artificial Intelligence

OCR - Optical Character Recognition

NLP - Natural Language Processing

API - Application Programming Interface

ML - Machine Learning

VPN - Virtual Private Network

GDPR - General Data Protection Regulation

PII - Personally Identifiable Information

E2EE - End-to-End Encryption

SDK - Software Development Kit

CPU - Central Processing Unit

GPU - Graphics Processing Unit

ML models - Machine Learning models

Resumen ejecutivo

Este documento se centra en el aprendizaje automático como servicio (MLaaS) y sus implicaciones en la privacidad. El documento hace hincapié en la necesidad de considerar las implicaciones de privacidad en todos los escenarios de MLaaS, ya que la información sensible puede inferirse a partir de datos aparentemente inocuos.

El documento explora diferentes escenarios de MLaaS, incluyendo ejemplos centrados en humanos y basados en IoT. Estos ejemplos demuestran los posibles beneficios del MLaaS, incluyendo el aprovechamiento de la alta potencia computacional para tareas complejas y la reducción de la necesidad de conocimientos especializados. Sin embargo, también destacan los posibles riesgos de privacidad asociados con la subcontratación de datos a un servicio de terceros.

El documento luego se adentra en escenarios de MLaaS sensibles a la privacidad con datos no estructurados, que pueden capturar y analizar inadvertidamente información sensible, lo que lleva a violaciones de privacidad. Estos escenarios demuestran la necesidad de medios técnicos y legales, como la anonimización de datos, el cifrado y las técnicas de ML que preservan la privacidad, para abordar las preocupaciones de privacidad en el MLaaS.

Finalmente, el documento concluye enfatizando la importancia de la educación y la conciencia del usuario para garantizar que las personas comprendan los riesgos de privacidad asociados con el MLaaS y tomen medidas adecuadas para proteger sus datos. También destaca la necesidad de investigación y desarrollo continuos de técnicas de ML que preservan la privacidad para permitir el uso seguro y responsable del MLaaS.

Abstract

This document focuses on machine learning as a service (MLaaS) and its privacy implications. The document emphasizes the need to consider privacy implications in all MLaaS scenarios, as sensitive information can be inferred from seemingly innocuous data.

The document explores different scenarios of MLaaS, including "human"-centered and IoT-based examples. These examples demonstrate the potential benefits of MLaaS, including leveraging high computational power for complex tasks and reducing the need for in-house expertise. However, they also highlight the potential privacy risks associated with outsourcing data to a third-party service.

The document then delves into privacy-sensitive MLaaS scenarios with unstructured data, which can inadvertently capture and analyze sensitive information, leading to privacy breaches. These scenarios demonstrate the need for technical and legal means, such as data anonymization, encryption, and privacy-preserving ML techniques to address privacy concerns in MLaaS.

Finally, the document concludes by emphasizing the importance of user education and awareness to ensure that individuals understand the privacy risks associated with MLaaS and take appropriate measures to protect their data. It also highlights the need for continued research and development of privacy-preserving ML techniques to enable the safe and responsible use of MLaaS.

1. Introduction

In recent years, major tech providers like Google, Microsoft, and Amazon have made it easier for customers to incorporate machine learning tasks into their applications through the use of software interfaces. This approach, known as Machine Learning as a Service (MLaaS), allows organizations to outsource complex tasks like training classifiers, performing predictions, and clustering. It also enables others to query models trained on their data. MLaaS is not limited to business partnerships, but can also be used in government collaborations and citizen science projects.

However, there are significant privacy concerns associated with MLaaS. If malicious actors can access the data used to train these models, it could result in severe information leakage. Moreover, if the inner parameters of the model are proprietary, then an adversary should not be able to access them. This paper reviews the privacy challenges in this space, analyzing the relevant research literature and exploring possible countermeasures.

The document provides a discussion on machine learning and privacy, discussing possible adversarial models and settings. It covers a discussion on the attacks that can lead to private and/or sensitive information leakage, and details techniquest attempting to defend against such attacks. Finally, it identifies open problems that require further attention.

2. A review on privacy concerns on Machine Learning

2.1. Machine Learning as a Service

Cloud providers, such as Microsoft, Amazon, and IBM, have introduced Machine Learning as a Service (MLaaS) offerings to enable clients to leverage machine learning capabilities without the expenses, time, and risks of developing in-house infrastructure. With MLaaS, small and medium-sized companies can access pre-built, generic machine learning tools such as predictive analytics, APIs, data visualization, and natural language processing, and customize them to their specific needs.

Purchasers of MLaaS services can use these tools through prediction APIs, and pay-per-query basis. Typically, the cost of an image classification service ranges from $1 to $10 per 1,000 queries, depending on the level of customization and sophistication of the machine learning model.

MLaaS services differ significantly among various providers. Some platforms allow clients to download and deploy machine learning models locally, while others only offer access to machine learning models through a prediction query interface, which provides the predicted label and confidence score. The latter is more popular. Additionally, some platforms permit clients to upload their own models and charge others for using them.

2.2. Privacy Enhancing Technologies (PET) for ML

Cryptography plays a vital role in protecting data confidentiality, specifically through encryption. In the context of machine learning and data analysis/processing, two essential cryptographic primitives are relevant: 1) secure multi-party computation[1] [2] (SMC), and 2) fully homomorphic encryption[3] [4] (FHE).

Secure multi-party computation enables two or more parties to collaboratively compute a function over their inputs while keeping their inputs hidden from each other. SMC protocols typically use tools such as garbled circuits, secret sharing, and oblivious transfer.

On the other hand, fully homomorphic encryption is an encryption scheme that allows processing of the underlying cleartext data while it remains in encrypted form, without giving away the secret key. In other words, FHE permits almost any computation over encrypted data.

Differential Privacy[5] (DP) addresses the challenge of learning useful information about a population without revealing anything about an individual. It provides statistical guarantees against what an adversary can infer from learning the result of a randomized algorithm. Differentially private techniques commonly protect individual data subjects' privacy by introducing random noise when producing statistics. In other words, DP guarantees that each individual is exposed to the same privacy risk, whether or not their data is included in a differentially private analysis

3. Target scenarios and use cases

In the following, we discuss the targeted scenario and some possible use cases tackled by these technologies.

3.1. Target scenario

In a typical non-privacy-preserving Machine Learning as a Service (MLaaS) scenario, there are two main entities: a Server (S) and a Client (C). C could be any client device with low computational power and/or memory available, controlled by either a human entity or a non-human one (IoT device). On the other hand, S possesses a pre-trained ML model deployed on her premises and offers a cloud service of ML classification via queries. C has some unlabeled data X and she wants predictions on it (Y). C sends X to S for processing, and S sends Y back to C, which marks the end of the typical MLaaS interaction.

Although this approach has some benefits, such as enabling C to leverage high-precision predictions from cloud machines with much higher computational power, it also has its drawbacks. One such issue is that once C sends X out on the public Internet, X becomes vulnerable to a variety of attacks.

While we assume that in-transit data integrity and confidentiality with respect to third parties are already granted by a simple SSL/TSL encryption, the real problem arises when the Server itself is assumed to be semi-honest. This means that S carries out her duty correctly and without altering the agreed-on protocol in any way (the prediction is indeed honestly computed and sent back unaltered, as requested). However, S could also decide to infer some unexpected, non-requested, and potentially sensitive information from X or sell the

[1] Mohassel, Payman, and Yupeng Zhang. "Secureml: A system for scalable privacy-preserving machine learning." 2017 IEEE symposium on security and privacy (SP). IEEE, 2017.

[2] Liu, Jian, et al. "Oblivious neural network predictions via minionn transformations." Proceedings of the 2017 ACM SIGSAC conference on computer and communications security. 2017.

[3] Gilad-Bachrach, Ran, et al. "Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy." International conference on machine learning. PMLR, 2016.

[4] Gentry, Craig. "Fully homomorphic encryption using ideal lattices." Proceedings of the forty-first annual ACM symposium on Theory of computing. 2009.

[5] Dwork, Cynthia, et al. "Calibrating noise to sensitivity in private data analysis." Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3. Springer Berlin Heidelberg, 2006.

unencrypted data to a third party that wants to do the same (Figure 1). After all, we assume that the data must be decrypted server-side sooner or later to allow processing (ML model prediction), unless Homomorphic Encryption is employed, which introduces other issues.

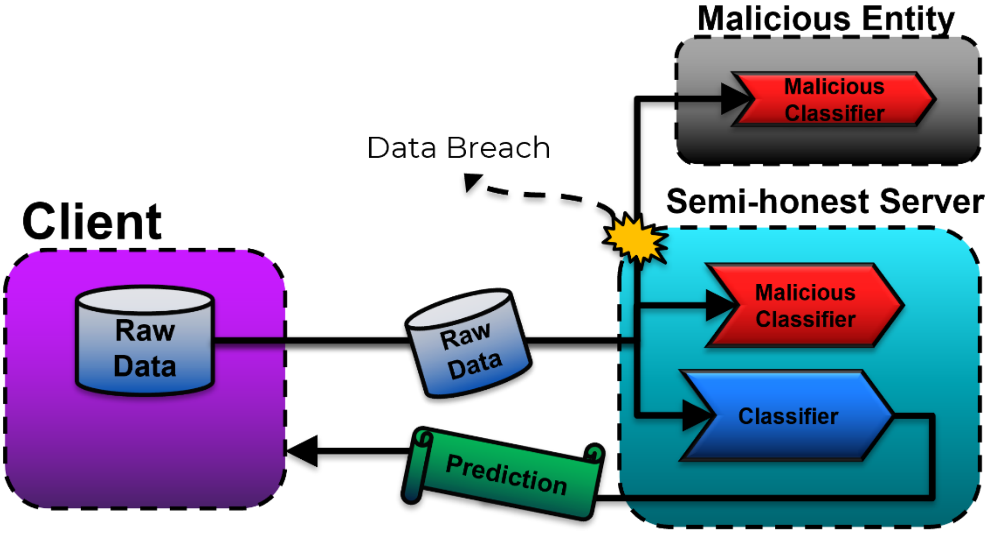

Figure 1: High-level view of a privacy related attack in a typical MLaaS scenario

3.2. Use Cases

Common scenarios for machine learning as a service (MLaaS) can be broadly classified into two categories: those centered around human interaction and those involving Internet of Things (IoT) devices.

Human-centered scenarios often involve applications that utilize various forms of human input such as images, speech, or natural language. Some examples of such applications include Google Lens/Photos for image classification and face recognition, digital assistants like Alexa/Siri for speech and sound recognition, Github Copilot for natural and programming language processing, and ChatGPT for content summarization and question answering using natural language processing.

On the other hand, IoT-centered scenarios generally involve sensor data collected by devices that are part of the IoT ecosystem. Examples of such devices include security cameras that utilize image classification and segmentation, face recognition, motion detection, and pattern recognition to provide enhanced security measures. Additionally, time-series data from various IoT sensors can be used for a variety of applications including predictive maintenance and anomaly detection.

Overall, MLaaS offers a broad range of applications and use cases for businesses and individuals to leverage machine learning capabilities without the need for costly infrastructure and expertise.

Examples of MLaaS scenarios that require privacy protection when dealing with unstructured data:

1. A mobile app that detects the mood of a driver via a continuous video stream sent to a cloud server. The ML model deployed on the server analyzes the data and sends commands to the app to issue a warning sound if the driver appears tired, sleepy, or distracted. The video data can potentially reveal sensitive information about the user, such as age, race, and gender, which could lead to biases and

privacy breaches.

2. Identity recognition systems that use facial scans for door control should ideally only distinguish between authorized and unauthorized individuals. However, the raw image data can be used for other purposes, such as identity recognition, gender/age/race/mood classification, without the users' knowledge or consent.
3. Crowd counting and people flux control systems that use video cameras can be made privacy-friendly by automatically blurring faces or deactivating anonymization in certain regions triggered by specific events, such as criminal activity recognition.
4. Physical activity recognition via smartphone sensors, such as the accelerometer/gyroscope, can reveal sensitive information, such as gender and medical conditions, unintentionally.
5. Voice-activated digital assistants, such as Alexa and Siri, can potentially capture sensitive information, such as credit card numbers, social security numbers, and personal matters, if they are unintentionally included in recorded commands.

4. Conclusion

In conclusion, the topics covered in this document have highlighted some of the important issues and challenges surrounding the use of Machine Learning as a Service (MLaaS) in various contexts. We have seen how MLaaS can offer a range of benefits, including the ability to leverage high precision predictions from cloud machines with much higher computational power, as well as the potential to improve the accuracy and efficiency of various applications in both human-centered and IoT scenarios.

However, we have also explored some of the risks and concerns associated with MLaaS, particularly in relation to privacy and security. The potential for data breaches, unauthorized access, and misuse of sensitive information is a significant issue that needs to be addressed in the development and deployment of MLaaS solutions. This is especially true in scenarios where unstructured data is involved, as there is often a wealth of extraneous information that can be inferred from such data, potentially leading to biases or privacy breaches.

Despite these challenges, it is clear that MLaaS has the potential to be a valuable tool in a range of applications, from facial recognition and security systems to digital assistants and people counting. As the technology continues to evolve and new methods for protecting privacy and security are developed, we can expect to see even more innovative and useful applications of MLaaS in the future. However, it is crucial that developers, businesses, and users work together to ensure that the benefits of MLaaS are balanced against the potential risks, and that appropriate safeguards are put in place to protect against unauthorized access, misuse, or abuse of sensitive data.