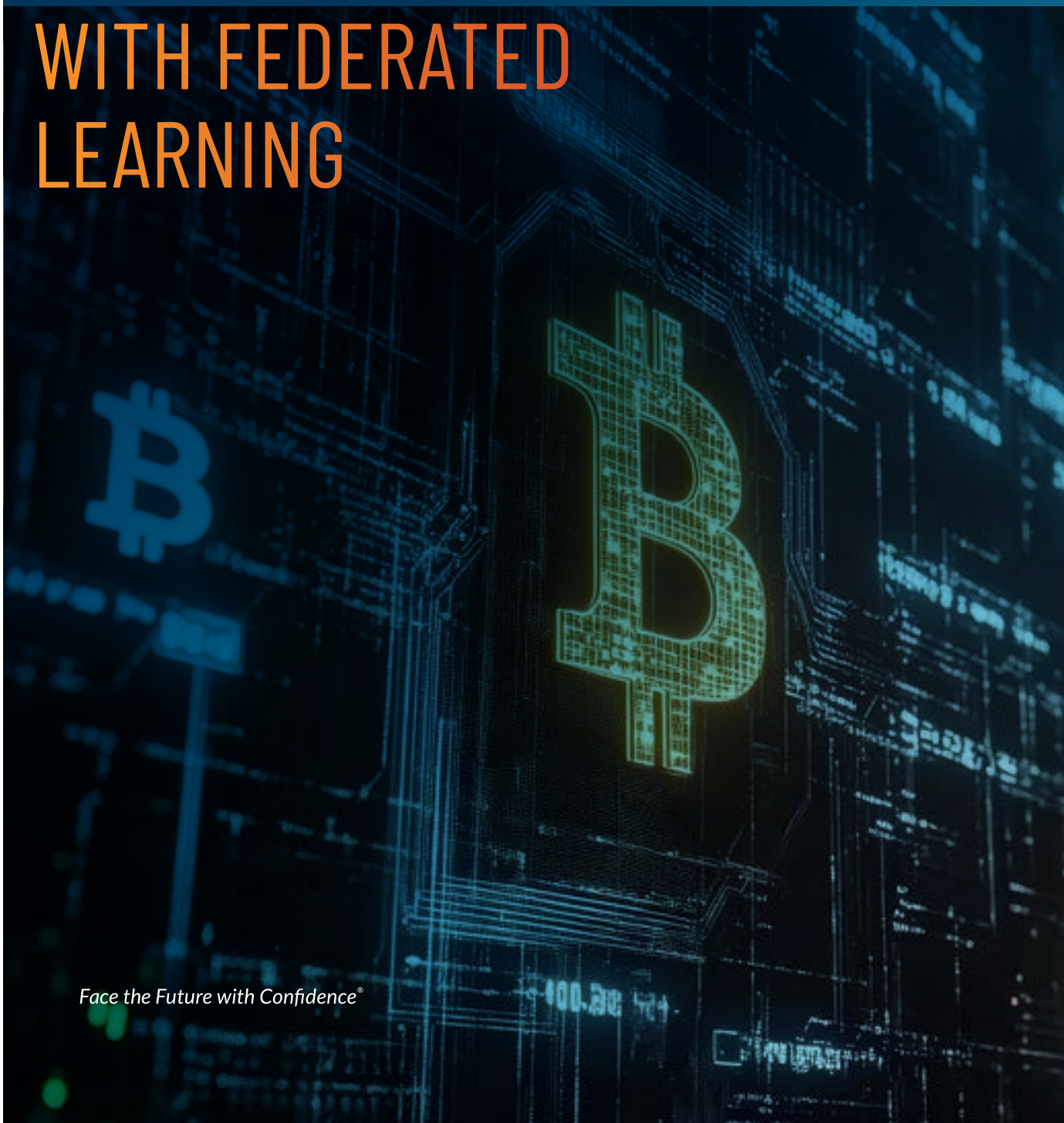
protiviti®
Global Business Consulting

HOMOMORPHIC ENCRYPTION INTEGRATED WITH FEDERATED LEARNING



Face the Future with Confidence®

TABLE OF CONTENTS

ABSTRACT

The whitepaper highlights the cryptographic technology which is Homomorphic Encryption (HE) integrated with Federated Learning (FL) while training machine learning (ML) models to provide a transformative approach to one of the biggest challenges faced by organizations in the current industrial scenario which is confidentiality of sensitive information and data privacy. Implementing FL can be achieved using multiple remotely based and decentralized local server models to train machine learning models while computing only the encrypted data from the server side.

The problem revolves around:

- Ensuring data privacy when training ML models since raw data is used at large which compromises sensitive data and integrity.
- Hindrance of trust, reducing productivity, violating protocols and regulatory non-compliance.
- Accuracy, scalability, and the reliability of the data which often plagues this process.

To overcome these challenges, we propose Federated Learning with Homomorphic Encryption:

- Federated Learning allows the machine learning model to use real-time data updates from the local models installed remotely while also maintaining data integrity and safeguarding sensitive information with the implementation of Homomorphic Encryption.
- HE facilitates curated utilization of the encrypted data to perform mathematical operations without compromising the data privacy or violating protocols when analyzing the data from datasets with similar features but different samples.
- This further helps in alleviating data breaches, foster trust, and attenuate risks.
- However, some of the limitations consist of higher computational overheads, bottlenecking on older GPUs.

This paper outlines the key components of the proposed Federated Learning (FL) model and highlights the advantages and effectiveness of the said model while keeping the data integrity intact and working with the cipher values to perform computations and train the ML model accordingly.

INTRODUCTION

In the rapidly advancing field of data science, the implementation of secure and efficient collaborative learning models is paramount for the success and competitiveness of businesses.

Federated Learning (FL) has emerged as a transformative approach, allowing multiple parties to train shared models without exchanging raw data, thereby preserving data privacy, and complying with stringent data protection regulations. However, the challenge lies in ensuring the confidentiality of the model updates, which can potentially leak sensitive information. Currently, the process of safeguarding these updates often depends heavily on traditional encryption methods and trust in the participating entities, introducing vulnerabilities and inconsistencies in data protection.

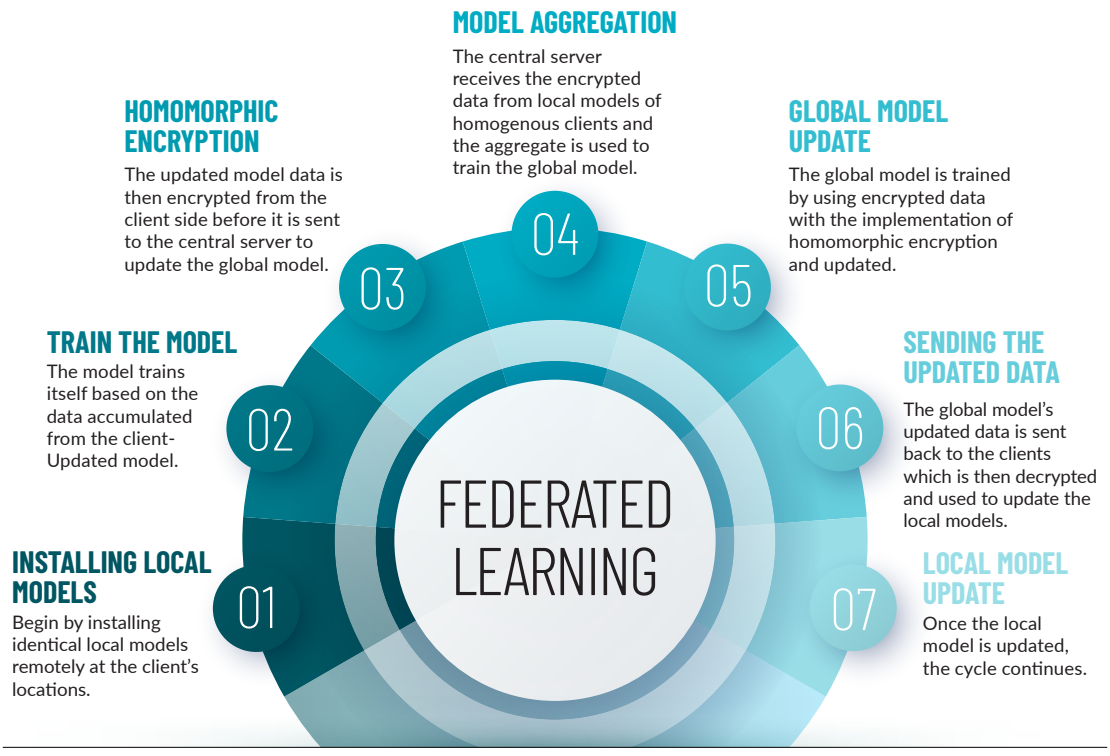
- Integration of **Homomorphic Encryption (HE)** into Federated Learning enables encrypted computations, ensuring confidentiality and integrity of model updates without decryption.
- Enhanced data security reduces the risk of unauthorized access and regulatory penalties, fostering a trustworthy environment for clients and compliance with data protection laws.
- Adoption of Federated Learning with Homomorphic Encryption facilitates seamless collaboration, sustained innovation, and confidence in building a reliable, secure foundation for future data-driven projects

1. PROBLEM STATEMENT

Banks and **E-Commerce** have always faced an issue when it comes to privacy of the consumers and subsequent data handling. The privacy preserving methods used by the sectors have never been optimal and consistent with the use of decrypted data at large by most firms. This puts sensitive data at risk from data breaches and cyber-attacks.

Homomorphic Encryption solves this problem by utilising the data from both ends in just the encrypted forms which enables data privacy and integrates it with already existing Federated Learning algorithms. When both these technologies are utilised, it can offer something for both fronts- banks and financial institutions. Namely personalised offers on bank's services on the commonly purchased products from the e-commerce platforms which will vary from consumer to consumer based on the frequency of purchase and the medium.

FEDERATED LEARNING



HOMOMORPHIC ENCRYPTION

- **Select a HE Scheme and Generate Keys**: Choose an appropriate homomorphic encryption scheme based on computational requirements and security needs, then generate the necessary cryptographic keys (public key, private key, and possibly evaluation keys).
- **Identify Data to Encrypt**: Determine the sensitive data that needs to be encrypted.
- **Apply Encryption:** Encrypt the identified data using the public key from the homomorphic encryption scheme.
- **Execute Encrypted Operations:** Perform the defined computations directly on the encrypted data using homomorphic operations (e.g., addition, multiplication).

2. EXISTING INDUSTRY IMPLEMENTATIONS

PROBLEMS

Data breaches have plagued all industries due to the boom of new technologies which extensively utilize online networks and cloud services. Healthcare industry is the one which gets the most affected by data breaches due to various factors:

- **Valuable Data:** Healthcare data, including patient records, is highly valuable on the black market. Personal Health Information (PHI) contains detailed personal, financial, and medical information that can be used for insurance fraud, identity theft and other malicious activities.
- **Legacy Systems:** Lot of healthcare organizations still use outdated software with hardware systems that are no longer supported or patched for security vulnerabilities. These legacy systems are easier targets for cyberattacks.
- **Third-Party Vendors:** Healthcare providers often collaborate with multiple third-party vendors for various services (e.g., billing, cloud storage, electronic health records systems). Each third-party connection is a potential entry point for attackers if not properly secured.

MARKET USE CASES

1. **NVIDIA collaborated with King's College London and French startup Owkin** to develop the first privacy-preserving federated learning system for medical image analysis.

Overview: Brain Tumour Segmentation using FL

- Utilized the **BraTS 2018 dataset** with pre-operative **MRI scans from 285 subjects**, including **four modalities** and annotations for whole tumor, tumor core, and enhancing tumor.
- Randomly split the dataset into a training set with **242 subjects** and a held-out test set with **43 subjects**. Stratified the training data into **13 subsets** based on the institution of origin and assigned each subset to a federated client to reflect real- world heterogeneity and imbalance.
- **Segnet5** was used with **Vgg16** for segmentation and image classification along with **hybrid CNN with Resnet18. Adam Optimiser** was used to evaluate model performance.
- **The centralized model**, where all data was available together, converged faster and required around **300 rounds** of training to achieve reliable performance levels. In contrast, **the federated learning model**, with data spread across different sources, needed approximately **600 rounds** of training to reach similar performance levels.
- The data-centralised model's training epoch with **242 data samples** took about **205 seconds**. However, for the federated model, the training time per epoch was dictated by the slowest data source, which had N=**77 samples**. Each epoch required around **65.45 seconds**, plus small overheads.

FOR HOMOMORPHIC ENCRYPTION

Each local server compiles the data for encryption and **differential privacy** was used with the **Clara v3.x, homomorphic encryption** later came with the newer **NVIDIA Clara train v4.0** which complies with both **PHI (Protected Health Information) and HIPAA (Health Insurance Portability and Accountability Act)**.

New Frontiers

USE CASE 1: PERSONALISED RECOMMENDATION SYSTEM FOR BANKS & E-COMMERCE

We suggest Vertical Federated Learning and HE for Banks and E- commerce platform:

- The idea is to use the data from ecommerce sites to analyse and find the most bought items and offer subsequent discount on them based on the data from the bank for the most common mode of purchase.
- This helps the banks with personalised financial services to the consumers as well as ecommerce platforms with services and sales, respectively.

STEPS

1. **Deploying local models with the banking and ecommerce app across devices-** Devices will include PCs, Mobiles for tracking usage real-time.
2. **Usage of Vertical Federated Learning for the bank and ecommerce platform-** Homomorphic encryption is used to convert the consumer data into cipher texts to maintain privacy and security.
3. Data to the global model for the bank updates itself with the aggregated data of the most used payment methods on the ecommerce platform.
4. Data to the global model for the ecommerce platform updates itself with the aggregated data of the most commonly purchases products. **(usage of Vertical FL- It utilizes various datasets with differing feature spaces to collaboratively train a global model based on same set of users)**
5. **Taking into account both the data sets-** Personalised offers based on the payment modes on the items can be offered to increase the rate of purchase- sales for the ecommerce website while also increasing the usage of the bank's services.

MARKET USE CASES

1. Bank

Protiviti suggests the FL model to be used is FATE open source, which comes with federated learning and can be used in the collection of data and analysis. It also provides a platform for utilising other open- source AI Models. Open-source capabilities also allow FATE (Federated AI Technology Enabler) with MPC (Multi party computation) to be used, which is a working example of FL in banking with its key implementation being WeBank's services. FATE also complies with GDPR and CCPA.

STEPS

1. Federated Recommendation Systems

Use Case:

- **Personalized Financial Products:** Recommending suitable payment methods, credit cards, or loan products based on usage patterns.

Key Techniques:

- **Collaborative Filtering:** Using user behaviour data to recommend payment methods that others with similar profiles prefer.
- **Matrix Factorization:** For personalized recommendations user-item interaction matrix is decomposed into latent factors.

Implementation:

- **Local Updates:** Train the recommendation model on local payment usage data.
- **Federated Averaging:** Aggregate the model updates to refine the global recommendation model.

Example:

- **Tencent, iOS keyboard, WeBank.**

2. Federated Clustering Models

Use Case:

- **Customer Segmentation: Segmenting customers based on payment methods usage for targeted marketing and personalized services.**

Key Techniques:

- **K-Means Clustering: Local K- Means models are trained on payment data to cluster customers based on their usage patterns.**
- **Hierarchical Clustering: Hierarchical clustering can be used to create nested clusters of customers.**

Implementation:

- **Local Clustering: Each institution clusters its customers based on payment method usage.**
- **Cluster Aggregation: Aggregated clustering results to form a global clustering model.**

E-COMMERCE

With industry experience and numerous projects, we have seen that real-time data processing helps with tracking customer searches, purchase history, and help train the local model. Seamless transaction monitoring helps both the bank and the platform since they are directly related. Marketing and sales analytics when it comes to products facilitates the ecommerce platform recommendations using data clustering.

STEPS

(User item interaction for search history and purchase history tracking)

1. Federated Collaborative Filtering

Overview:

- Federated Collaborative Filtering (FCF) leverages user interactions (such as purchase and search history) across multiple devices to improve recommendation systems.

Key Techniques:

- **Matrix Factorization:** For personalized recommendations user-item interaction matrix is decomposed into latent factors.
- **Neural Collaborative Filtering (NCF):** Utilizes neural networks to model complex user-item interactions.

Implementation:

- **Local Training:** Each user's device trains the model on their own interaction data.
- **Federated Averaging**
- **(FedAvg):** Without sharing raw data, aggregates the model updates from multiple devices to create a global model.

Example:

- **Google's Federated**
- **Recommender:** Google uses FCF for apps like Google Play, where local interactions are used to improve app recommendations. **WeBank**
- **Financial Services, GBoard**.

FOR HOMOMORPHIC ENCRYPTION ON BOTH ENDS

We suggest Paillier algorithm-based PHE (Partial Homomorphic Encryption) by utilising the multi-buffer functions like Intel® Integrated Performance Primitives Cryptography (Intel® IPP- Cryptography) library which is open source. Same has been utilised before in case of WeBank when it partnered with Intel in order to accelerate the modular exponentiation operation of PHE.

The general workflow HE protocol- based Federated Learning-

- **Key:** Generation of Public key by the arbiter and distribution to each party.
- **Encryption:** Original data of each party is encrypted and the uploaded to the federated learning environment.
- **Operation:** To use encrypted data in a federated learning environment to carry out activities (like model training)
- **Decryption:** The resulting optimization model is returned to participants, and then decrypted.

By default, the FATE framework adopts the GNU MP Bignum Library (GMP) which is an open-source mathematical operation library, in order to perform the modular exponentiation operation. The GMP library still uses "serial" computing requests though it can support arbitrary-precision mathematical operations, even when the length of the data involved in the operation reaches thousands of bits. Newer hardware helps with faster computations.

Limitations and possibilities:

Bottlenecking is the main limitation when using PHE (Partial Homomorphic encryption) which limits computation speeds of the encrypted data in the solutions.

There are libraries currently in development which are actively working towards addressing the bottlenecking.

The most prominent one being:

Intel HE Acceleration Library which works on inverse negacyclic Number- Theoretic Transform (NTT), element-wise vector, vector modular operations. Intel AVX-512 IFMA Instruction Set is being used to develop the library and fine tune it for more efficiency.

USE CASE 2: ENHANCING TELECOM NETWORK PERFORMANCE WITH FEDERATED LEARNING AND HOMOMORPHIC ENCRYPTION

01
LOCAL DATA COLLECTION
Description: Each network node (e.g., base stations, user devices) collects data on traffic load, latency, bandwidth usage, etc.

02
LOCAL MODEL TRAINING
Description: Using federated learning, each node trains a local machine learning model on its own data.

03
ENCRYPTED MODEL UPDATES SENT TO CENTRAL SERVER
Description: The locally trained model parameters are encrypted using homomorphic encryption before being sent to the central server.

04
MODEL AGGREGATION
Description: The central server aggregates the encrypted model updates from all nodes, using homomorphic encryption to combine them without decryption.

05
GLOBAL MODEL UPDATE
Description: The aggregated global model is updated and redistributed to all local nodes.

06
LOCAL MODEL APPLICATION
Description: Each node applies the updated global model to optimize routing, predict and prevent congestion, and enhance QoS.

07
ENCRYPTED PERFORMANCE METRICS COLLECTION
Description: Nodes encrypt their performance metrics (e.g., latency, packet loss) and send them to the central server.

08
SECURE ANALYSIS AT CENTRAL SERVER
Description: The central server performs analysis on the encrypted performance metrics using homomorphic encryption.

09
ENCRYPTED CONFIGURATION ADJUSTMENTS SENT BACK TO NODES
Description: The central server sends encrypted recommendations or configuration adjustments back to the local nodes.

10
LOCAL DECRYPTION AND APPLICATION OF ADJUSTMENTS
Description: Local nodes decrypt the received adjustments and apply the necessary changes to optimize network performance.

Limitations:

- Variations in data collected from different nodes may limit model effectiveness.
- Homomorphic encryption introduces computational overhead, impacting real-time processing.
- Delays in aggregating and redistributing global models may affect real-time responsiveness.
- Effective implementation relies on cooperation and adoption by all network operators.

FEDERATED LEARNING MODEL

Model: Federated Averaging (FedAvg)

Overview:
This technique enables telecom networks to improve model accuracy and performance while respecting data privacy.

STEPS

Local Model Training:

- Each network node trains a local machine learning model on its own data, capturing insights specific to its operational environment.
- Training can utilize various models such as neural networks (CNNs, RNNs), decision trees, or other suitable algorithms.

Model Aggregation:

- Encrypted model updates (weights) from local models are transmitted to a central server.
- The central server aggregates these updates using federated averaging or another aggregation method to create a global model.

Global Model Distribution:

- The updated global model is redistributed to all nodes, providing each with enhanced insights derived from the entire network's data.
- Nodes integrate these updates to refine their local models and improve decision-making capabilities.

Advantages:

- **Data Privacy:** Raw data remains decentralized and never leaves the local nodes, ensuring privacy and compliance with regulations.
- **Collaborative Improvement:** Collective learning from diverse data sources enhances model robustness and adaptability to varying network conditions.
- **Reduced Data Transfer:** Only model updates are transmitted, reducing bandwidth usage compared to transmitting raw data.

HOMOMORPHIC ENCRYPTION

Model: CKKS (Cheon-Kim-Kim-Song) Scheme:

Overview:
In telecom networks, it enables secure aggregation and analysis of sensitive data such as model updates and performance metrics while preserving confidentiality.

STEPS

Data Encryption:

- Model updates, performance metrics (e.g., latency, packet loss), or other sensitive data are encrypted using a homomorphic encryption scheme like Paillier.

Encrypted Computation:

- Encrypted data undergoes computation (e.g., aggregation, analysis) directly on the ciphertext without decryption, leveraging homomorphic properties.

Secure Transmission and Storage:

- Encrypted results or recommendations are securely transmitted and stored, maintaining data confidentiality throughout the process.

Advantages:

- **Data Confidentiality:** Protects sensitive information from unauthorized access or exposure during transmission and processing.
- **Compliance:** Facilitates compliance with data protection regulations by safeguarding data privacy.
- **Scalability:** Supports scalable and secure data processing across distributed networks without compromising security.

Challenges:

- **Computational Overhead:** Homomorphic encryption can be computationally intensive, impacting processing speed and resource usage.
- **Algorithmic Limitations:** Current homomorphic encryption schemes may have limitations on the types of computations or operations that can be efficiently performed.

COMPANIES IMPLEMENTING FEDERATED LEARNING AND HOMOMORPHIC ENCRYPTION IN TELECOM NETWORKS

Nokia:

- Nokia Bell Labs has explored federated learning for telecom networks to enhance network performance while maintaining data privacy.

Ericsson:

- Ericsson is working on integrating AI and machine learning into telecom networks, including federated learning techniques to optimize network performance while safeguarding user data.

3. FUTURE DIRECTIONS

The future of federated learning in the age of homomorphic encryption lies in enhancing privacy and security protocols while also maintaining regulatory compliance. through advanced encryption schemes and algorithmic efficiency. Practical applications in healthcare as seen in the case of NVIDIA and King's College London and in case of banking- WeBank, highlight its potential.

Interoperability will further drive widespread adoption and trust, revolutionizing collaborative machine learning by ensuring data privacy and security which has been the main concern of the industries while opting for machine learning based solutions.

Both of these technologies combined facilitate in fostering ideas and implementations in new sectors as well as improvements in existing sectors along with integrations from third parties as seen in the suggestive use cases.

4. CONCLUSION

In conclusion, the integration of federated learning with homomorphic encryption represents a much-needed advancement in privacy-preserving machine learning, enabling secure and confidential collaborative model training. This combination addresses critical data privacy and security concerns by allowing computations on encrypted data, ensuring that individual contributions remain protected throughout the process.

While challenges such as computational, operational overheads and communication efficiency persist, ongoing research is steadily enhancing the efficiency of these techniques. This aligns with the growing emphasis on ethical AI and data protection, promising a future where diverse datasets can be leveraged for innovative AI solutions without compromising individual privacy.

5. REFERENCES

1. Roth, H. (2022, September 2). Federated Learning with Homomorphic Encryption NVIDIA Technical Blog. NVIDIA Technical Blog. https://developer.nvidia.com/blog/federated-learning-with-homomorphic-encryption/
2. Alarcon, N. (2022, September 2). NVIDIA and King's College London Debut First Privacy-Preserving Federated Learning System for Medical Imaging | NVIDIA Technical Blog. NVIDIA Technical Blog. https://developer.nvidia.com/blog/first- privacy-preserving-federated-learning-system/
3. NVIDIA Clara Train 4.0 - NVIDIA Docs. (n.d.). NVIDIA Docs. https://docs.nvidia.com/clara/clara-train-archive/4.0/index.html
4. Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M. J., & Feng, A. (2019, October 2). Privacy-preserving federated brain tumour segmentation. arXiv.org. https://arxiv.org/abs/1910.00962
5. Accelerating secure compute for federated Learning on the FATE. . . (n.d.). Intel. https://www.intel.com/content/www/us/en/developer/articles/technical/homomorphic- encryption/accelerating-secure-compute-for-fate-framework.html
6. Repository of Use Cases | PETs Adoption Guide. (n.d.). PETs Adoption Guide. https://cdeiuk.github.io/pets-adoption-guide/repository/
7. Corporation, N. (2024, February 26). Nokia launches blockchain-powered Data Marketplace for secure data trading and AI models. Nokia. https://www.nokia.com/about-us/news/releases/2021/05/05/nokia-launches- blockchain-powered-data-marketplace-for-secure-data-trading-and-ai-models/
8. Khan, M. J., Fang, B., & Zhao, D. (2023, September 13). Toward lossless homomorphic encryption for scientific computation. arXiv.org. https://arxiv.org/abs/2309.07284

ABOUT PROTIVITI

Protiviti (www.protiviti.com) is a global consulting firm that delivers deep expertise, objective insights, a tailored approach and unparalleled collaboration to help leaders confidently face the future. Protiviti and its independent and locally owned member firms provide clients with consulting and managed solutions in finance, technology, operations, data, digital, legal, HR, risk and internal audit through a network of more than 90 offices in over 25 countries.

Named to the 2024 Fortune 100 Best Companies to Work For® list for the past 10 years, Protiviti has served more than 80 percent of Fortune 100 and nearly 80 percent of Fortune 500 companies. The firm also works with government agencies and smaller, growing companies, including those looking to go public. Protiviti is a wholly owned subsidiary of Robert Half Inc. (NYSE: RHI). Founded in 1948, Robert Half is a member of the S&P 500 index.

Contact us

Amit Lundia
Managing Director
Phone: +91 33 66571501
Mobile: +91 9836922881
Email: amit.lundia@protivitiglobal.in

Kallol Kumar
Director
Phone: +91 33 66571510
Mobile: +91 9830087931
Email: kallol.kumar@protivitiglobal.in

Acknowledgement

Chiranjib Sarma, Associate Director from our Technology & Digital Solution Practice has contributed to this publication, which was led by Amit Lundia.

Face the Future with Confidence®

protiviti

PRO_DS_93728_FEB25