Slovak University of Technology
Faculty of Informatics and Information Technologies

Ing. Oleksandr Lytvyn

Dissertation Thesis Abstract

Machine Learning and Sensitive Data Protection

to obtain the Academic Title of
"philosophiae doctor", abbreviated as "PhD."

in the doctorate degree study programme:
9.2.9. Applied Informatics

in the field of study:
18. Computer Science

Form of Study:
Full-Time

Place and Date:
Bratislava, 26.08.2025

STU

Dissertation Thesis has been prepared at:
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava, Slovakia

Submitter: Ing. Oleksandr Lytvyn
Institute of Informatics, Information Systems and Software Engineering
Faculty of Informatics and Information Technologies,
Slovak University of Technology
Ilkovičova 2, 842 16 Bratislava

Supervisor: doc. Ing. Giang Nguyen Thu, PhD.
Institute of Informatics, Information Systems and Software Engineering
Faculty of Informatics and Information Technologies,
Slovak University of Technology
Ilkovičova 2, 842 16 Bratislava

Dissertation Thesis Abstract was sent: .....

Dissertation Thesis Defence will be held on: .....

At: ..... (am/pm) **at:** .....

prof. Ing. Ivan Kotuliak, PhD.
Dean of the Faculty of Informatics and Information Technologies

Abstract:

Sensitive information is central to many critical sectors, enabling machine learning approaches to derive valuable insights, yet the distributed and private nature limits its full utilization. Organizational silos, legal regulations, and the absence of trust between entities, require privacy-preserving alternatives to direct data sharing. These challenges are especially pronounced in cybersecurity, where unauthorized data access could expose critical vulnerabilities and amplify system threats. Therefore, this thesis explores the application of Privacy Enhancing Technologies to secure collaboration on sensitive data. The hybrid privacy-preserving learning approach is proposed for distributed network monitoring. The approach uses Federated Learning for collaborative learning through the exchange of model weights instead of raw data. To enhance security during aggregation, Secure Multi-Party Computation is employed, ensuring that shared artifacts remain protected. The experimental results demonstrate the approach's feasibility in multiple data distribution settings, with the resulting models outperforming their single-client counterparts, validating the effectiveness of collaborative privacy-preserving training. The proposed solution fills the gap in privacy-preserving collaboration in the cybersecurity space, allowing multiple entities secure machine learning collaboration without compromising data privacy or requiring mutual trust.

Abstract in Slovak Language:

Citlivé informácie sú kľúčové pre mnohé kritické odvetvia, pretože umožňujú strojovému učenie odvodiť cenné poznatky. Avšak ich distribuovaný a dôverný charakter obmedzuje ich plné využitie. Organizačné bariéry, právne predpisy a nedôvera medzi subjektami si vyžadujú súkromie chrániace alternatívy priameho zdieľania údajov. Tieto výzvy sú obzvlášť výrazné v kybernetickej bezpečnosti, kde neoprávnený prístup k údajom môže odhaliť kritické zraniteľnosti a zhoršiť hrozby pre systémy.Preto táto práca skúma aplikáciu technológií na zlepšenie ochrany súkromia na zabezpečenú spoluprácu s citlivými údajmi. Navrhuje sa hybridný prístup k ochrane súkromia pri distribuovanom monitorovaní sietí. Tento prístup využíva federatívne učenie na kolaboratívne učenie výmenou váh modelov namiesto surových údajov. Na zvýšenie bezpečnosti počas agregácie sa používa zabezpečený výpočet viacerých strán čo zaisťuje ochranu zdieľaných artefaktov.Experimentálne výsledky dokazujú realizovateľnosť tohto prístupu v rôznych distribučných nastaveniach, pričom výsledné modely prekonávajú modely trénované jednotlivými klientmi, čo potvrdzuje efektivitu kolaboratívneho učenia s ochranou súkromia. Navrhované riešenie zaplňuje medzeru v spolupráci s ochranou súkromia v kybernetickej bezpečnosti, umožňujúc viacerým subjektom bezpečnú spoluprácu v strojovom učení bez ohrozenia dátového súkromia alebo potreby vzájomnej dôvery.

Contents

Chapter 1

Introduction

Sensitive data is a fundamental component of operations across industries such as cybersecurity, healthcare, and finance. Machine Learning (ML) serves as a powerful tool for extracting insights and patterns from sensitive data, enabling more informed decision-making. However, the decentralized and private nature of sensitive data makes direct access and centralized learning infeasible. Such circumstances emerge the collaborative learning trend without compromising data privacy. Developing effective intelligent solutions requires diverse and consistent data, which is often lacking within individual organizations. Cross-organizational data sharing can enhance model performance but is hindered by legal barriers (e.g., GDPR [1], DSA [2], AI Act[3]), organizational silos, and a lack of trust between entities.

Privacy-Enhancing Technologies (PETs) have emerged to address this challenge, with approaches such as Federated Learning [4], Secure Multi-Party Computation [5], and Differential Privacy [6] with high potential in enabling ML on distributed sensitive data. While many existing approaches focus on individual PETs, their integration remains a significant challenge. A hybrid approach that ensures both privacy and efficiency is critical for practical deployment.

This work therefore focuses on developing a hybrid privacy-preserving learning framework for collaborative threat detection across organizations. To achieve this, PETs are examined and compared based on privacy, performance, and applicability across the ML lifecycle. Next, the work outlines the hybrid approach's design and system architecture, followed by evaluation in simulated real-world scenarios. The proposed approach addresses a gap in enabling privacy-preserving collaboration in cybersecurity and other sensitive domains, providing a practical solution for collaboration between organizations for mutual benefit without establishing strict mutual trust.

Chapter 2

Sensitive Data Protection in Collaborative Machine Learning

Gathering data in one place becomes a more complicated task because of increasing size together with strict privacy and security regulations. On the other hand, the utilization of distributed data maintains its demand, encouraging the development of collaborative learning approaches. Here the FL is discussed as the most prominent and approach for collaborative learning with inherent privacy-preserving capabilities.

2.1 Federated Learning

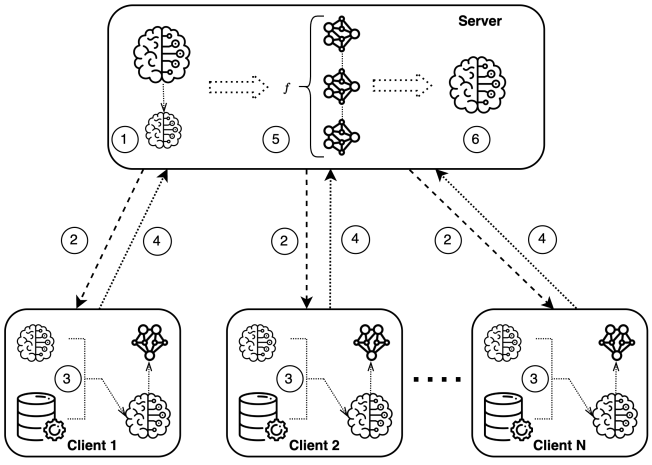


Figure 2.1: Federated Learning Architecture

Federated Learning (FL) is a distributed ML approach that allows multiple parties to collaboratively train a machine learning model while keeping their data private.

The process of FL training can be outlined as follows: (1) The server provides model snapshots to the clients, or sometimes clients agree on what model is to be trained. (2) Each client receives a copy of the model or the weights that define it; (3) The clients then train the model locally using their data; (4) After training, the clients send the updated model weights or the changes relative to the initial model back to the server; (5) The server aggregates all weights based on the chosen aggregation strategy; (6) The server applies weights to the model, and evaluates the model. Subsequently, the server distributes the updated model to all clients, and the process repeats from step (2) for a predetermined number of rounds until model convergence is achieved [7].

Yang et al. [8] categorize FL architectures into two primary types based on data structure: Horizontal Federated Learning (HFL), applicable when datasets share feature space but have distinct samples, and Vertical Federated Learning (VFL), employed when datasets share sample IDs but possess differing feature spaces.

2.2 Privacy Enhancing Technologies

Differential Privacy DP is one of the approaches to preserve privacy using a statistical data perturbation [6].

Definition 1. *A randomized algorithm K with domain $\mathcal{D}$ is (ε, δ)-differential private if $\forall S \subseteq$ Range(K) and $\forall a, b \in \mathcal{D}$ such that $\|a - b\|_1 \leq 1$ (a and b are adjacent inputs), if the following condition is valid.*

$$Pr[K(a) \in S] \leq exp(\varepsilon) \times Pr[K(b) \in S] + \delta \tag{2.1}$$

where K is a randomized algorithm; S is the set of possible outcomes of K; epsilon (ε) is the maximum distance between $K(x)$ and $K(y)$, then $\varepsilon \geq 0$ (also referred to as the privacy budget); delta (δ) is the information leakage probability, then $\delta \in [0, 1]$; Pr is the probability taken over the coin tosses of K.

(ε, δ)-*DP* ensures that the removal of any individual participant from the dataset does not significantly alter the output distribution. Conversely, (ε, δ)-*DP* provides formal guarantees by limiting information disclosure through constrained data release mechanisms, such as calibrated noise addition to query results [9].

Secure Multi-Party Computation Secure Multi-Party Computation (SMPC) is a generic cryptographic framework that allows multiple parties without mutual trust to collaboratively compute a function while keeping their respective inputs private. SMPC can operate under multiple security models that define the level of robustness against malicious adversary behavior. Here belongs Semi-honest , Malicious, and Covert adversary models [10].

There multiple techniques through which SMPC can be implemented, including Secret Sharing (SS) [5], Private Set Intersection (PSI) [11], Threshold Homomorphic Encryption (THE) [12], Oblivious Transfer (OT) [13], and Zero-Knowledge Proof (ZKP) [14].

Narrowing the scope to protocols suitable for ML operations, the number of suitable solutions decreases significantly. SMPC in ML context often considers only a small number of participants, from 2 to 4.

One of the first protocols that is efficient and suitable for ML is Function Secret Sharing (FSS) based on splitting and secret sharing of the function, while maintaining the possibility of correct computation is still [15]. ABY^3 is another SMPC protocol for three-party computation based on arithmetic sharing, Boolean sharing, and Yao's garbled circuits that implements semi-honest and malicious security schemes and is suitable for DNN evaluation [16]. Falcon is a recent Secret Sharing-based SMPC protocol designed for secure DNN evaluation, optimized for 3-party computation in both malicious and semi-honest settings. It supports essential DNN operations like linear, convolution, ReLu, Maxpooling, and batch normalization layers [17].

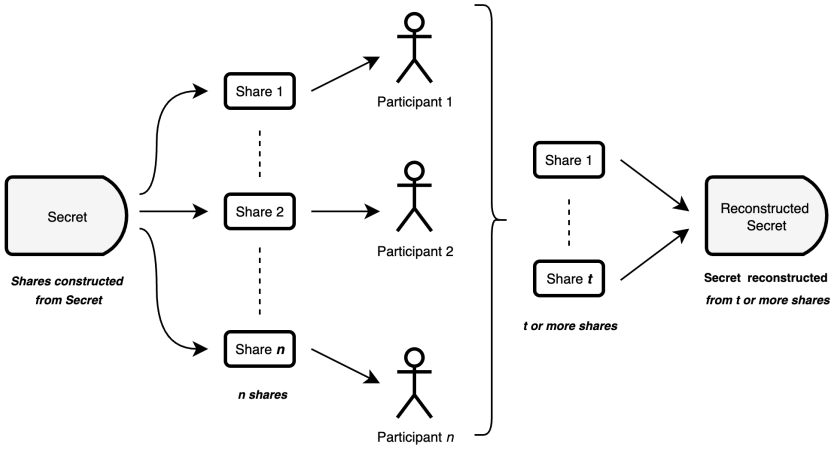


Figure 2.2: Secret Sharing process under (t, n)-threshold scheme

Secure Aggregation in Federated Learning Federated Aggregation (FA) is the process of aggregating artifacts received from participants, what are the results of one round training. Secure Aggregation (SA) is a form of FA focused on protecting the privacy of individual client updates during global model aggregation. While client updates does not contain raw data, they remain vulnerable to attacks such as membership inference [18], reconstruction [19], property inference [20], and model extraction [21]. Most FL approaches address this by applying only aggregated updates to the global model, preventing direct access to individual client updates during the aggregation process. One of the first protocols for implementing secure aggregation in the context of FL is SecAgg. Except for Secret Sharing, the SecAgg protocol utilizes the following cryptographic primitives: Key Agreement, Authenticated Encryption, Pseudorandom Generator, Signature Scheme, and Public Key Infrastructure [22].

Chapter 3

Hybrid Design for Sensitive Data Protection in Federated Learning

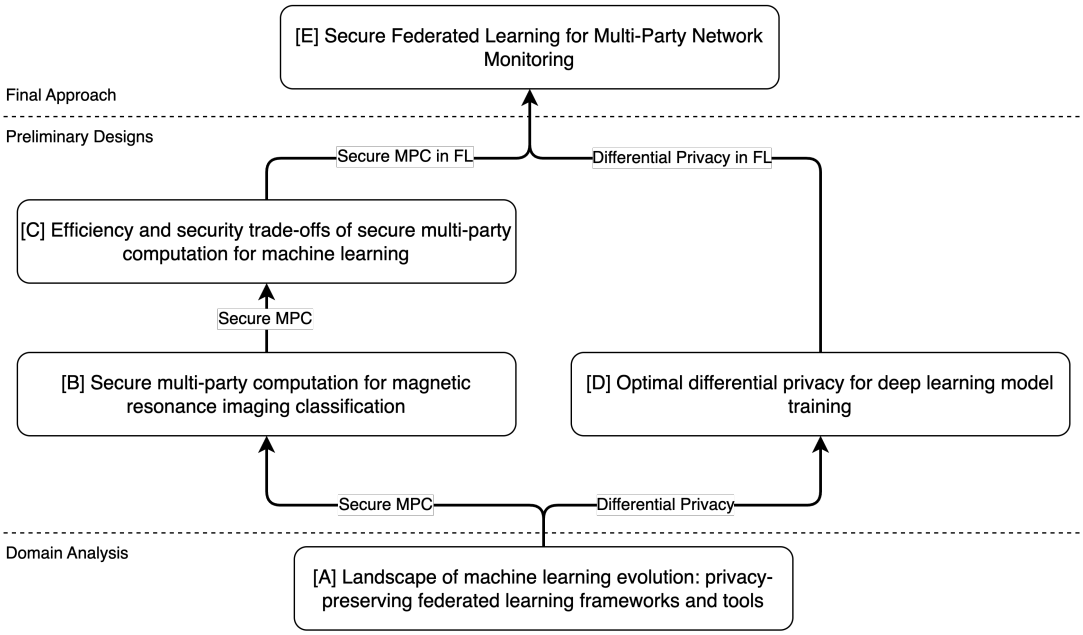


Figure 3.1: Research trajectory from the publications standpoint

Domain analysis highlighted the existence of various approaches to secure and preserve the privacy of sensitive data within the collaborative machine learning life-cycle. The conceptual design of the PETs application positions sensitive data at its core. This central placement reflects the need to consider diverse types of sensitive information, each representing a potential input for machine learning processes.

Further sections follow the structure outlined in Figure 3.1. The first publication [A] examines the current state of the domain and analyzes existing approaches. Following the domain analysis, preliminary designs for the PETs application are developed within a machine learning context (publications [B,C,D]). Considering the preliminary designs and experiment results, the final approach to collaborative machine learning (publication [E]) is proposed and discussed in Section 3.2.

3.1 Preliminary Designs Findings

Publication [B] focuses on the assessing the functionality of SMPC usage within collaborative ML for medical image classification. It proves that SMPC can be used for the joint inference. Yet, SMPC comes with the price of increased computation and communication loads, making the practical usage infeasible. This finding suggests that future work should explore the capabilities of SMPC protocols in greater depth.

Publication [C] explores the capabilities and efficiency of available SMPC protocols. The results indicate that while model performance decline is minimal, the computational load has significant variance depending on the adversary model and the number of participants. Thus emphasizing the importance of security and efficiency requirements definition for a specific use case to select appropriate protocol and settings. Another outcome of the case study is a new perspective for SMPC usage within ML context, where computation requirements are lower and the privacy of underlying information matters. This shifts the research perspective towards the Federated Learning, where only one point of direct sensitive information exposure is present - weights aggregation of participating clients.

Publication [D] adopts a different approach by exploring Differential Privacy to preserve data privacy during local training while minimizing model performance degradation. Main thought behind is that application DP-enabled local training for federated learning could provide an alternative to SMPC in protection of shared weights against unauthorized access or other exposure. The core finding is the existence of a relationship between regular and privacy parameters. Specifically, DP parameters influence the final model performance, but it is possible to achieve non-private model metrics through tuning mechanisms. In the Federated Learning (FL) context, this showcases the ability of DP to protect shared weights, without compromising the model's performance.

The findings outlined reveal two key circumstances:

- First, while SMPC represents a viable option for ML inference, its current implementations remain impractical for real-world applications due to computational intensity, pushing us toward less demanding use cases within the collaborative ML lifecycle.

- Second, differential privacy (DP) demonstrates comparable performance to non-private counterparts, though achieving this requires problem-specific optimization.

These case studies formulate preliminary designs phase and establish the context for the subsequent chapter.

3.2 Secure Federated Learning for Multi-Party Network Monitoring

This section introduces the design of a collaborative and privacy-preserving machine learning system, as presented in publication [E] of the research trajectory in Figure 3.1.

Modern network security relies on continuous monitoring through SOCs using IDS/IPS and behavioral analysis (NBA), requiring adaptive approaches for evolving threats. Machine learning enables predictive threat detection, but data limitations and privacy regulations like GDPR restrict model training. This work explores Federated Learning with secure multi-party computation (SMPC) to enable collaborative, privacy-preserving threat detection across distributed SOCs—balancing improved threat coverage against increased computational costs and privacy risks from multi-party participation.

The core motivation of this section is assessing the multi-party system behavior in the model-sharing training process, where the dynamic workload modeling is conducted in a secure and privacy-preserving manner among distributed and collaborated SOCs. To achieve this, our scope is to cooperate the Federated Learning (FL) concept and time-series prediction using DL techniques, along with the secure aggregation by means of secret sharing in FL. Concretely, SMPC in federated network monitoring, where more parties train the shared model over private data without being exposed to potential data privacy and data security threats.

3.2.1 Conceptualization of Secure FL for Network Monitoring

The high-level architecture of FL is visualized in Figure 3.2, which contains N clients and an orchestrating server server. The essential unit of FL training is *round*, which represents a sequence of next steps:

1. Each client trains model locally, usually for one epoch;
2. Each client sends model weights to the server;
3. Server aggregates weights from all clients;
4. Each client receives the update and starts a new cycle;

Each of these steps, represents a set of challenges. However, the most challenging step is step three, which requires an optimal approach to aggregate weights that would

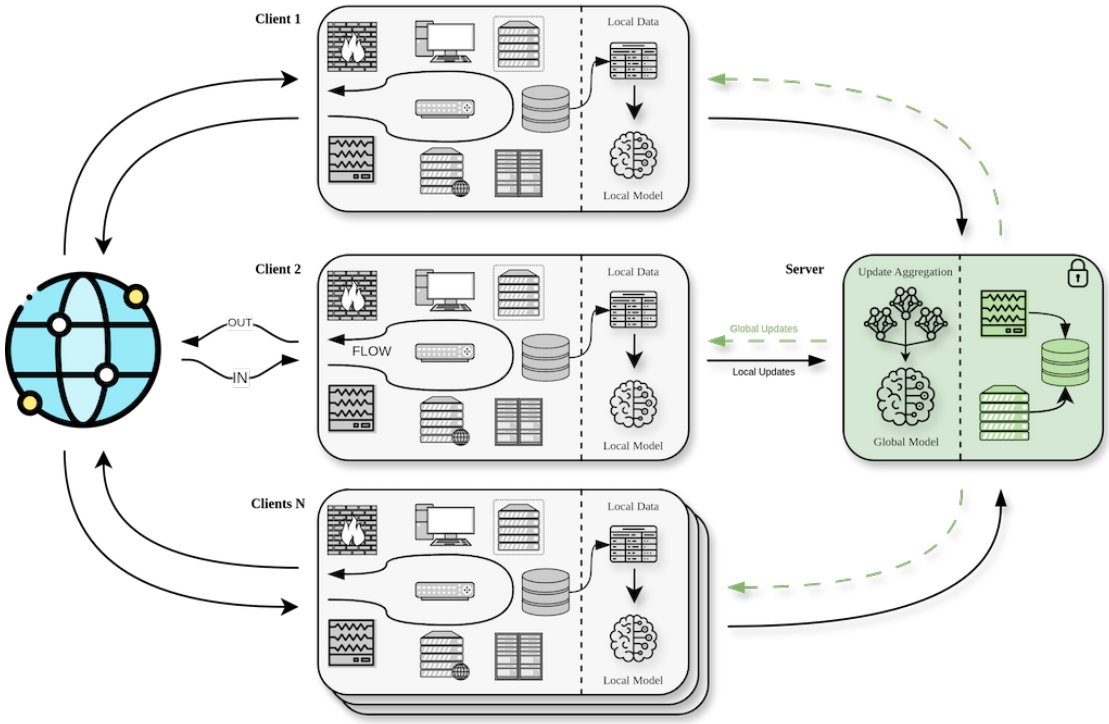


Figure 3.2: Conceptualization of Federated Network Monitoring System

consider the amount and distribution of data on each client, possible dropout clients, and other possible issues.

The federated network monitoring process is divided into several stages as follows.

- Stage 1 – data collection and data preparation, where all data of network activity have been collected and processed [23] to be more suitable for time series modeling.
- Stage 2 – design and creation of the DL model [24].
- Stage 3 – includes setting up the baseline experiments and model architecture optimization. After that, the best model is trained on different data partitions with subsequent evaluations.
- Stage 4 – establishes the FL process and training on the initial number of clients on all data partitions.
- Stage 5 – adds secure weight aggregation on the central server to the settings in the previous part.
- Stage 6 – collects metrics such as model performance metrics, prediction results, convergence rates, and computational loads are collected, compared, and evaluated appropriately.

3.2.2 System Architecture

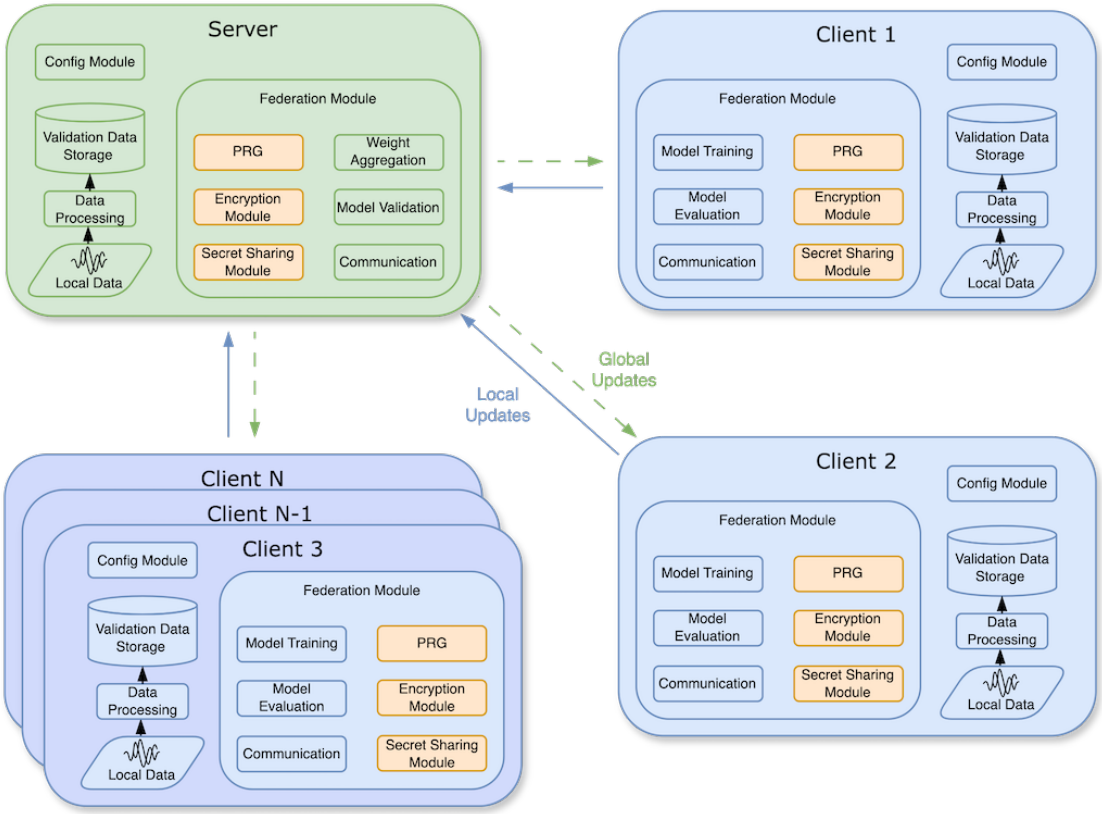


Figure 3.3: Federated System Architecture for Federated Network Monitoring

The system design is built based on the FL approach (Section 2.1) and visualized in Figure 3.2. Each client possesses a similar set of components which can be modified according to the chosen aggregation approach. For example, the implementation of Secure Aggregation requires additional modules on the client and server sides. In this section, the system components and their inter-relationships are discussed.

In a standard and non-secure setting, the client's component set comprises Data Processing, Data Storage, Configuration Module, Federation Modules, which contains submodules Model Training, Model Evaluation, and Communication.

In Secure Aggregation, additional components like Pseudo-Random Generator (PRG), Secret Sharing Module, and Encryption modules are added. The server architecture mirrors the client-side structure but incorporates three distinct core modules - Weight Aggregation, Validation Data Processing and Storage, and Model Validation - with adapted recurring modules to handle orchestration tasks, while supporting both horizontal and vertical federated learning data partition approaches.

3.2.3 Federated Process

The joint ML model training process in this work follows the standard methodology in the FL domain. The central server coordinates the exchange of information and oversees the aggregation of model weights, while each client trains their respective models independently on local systems.

We consider common prerequisites such as the existence of the Server and Set of Client Nodes with the ability to communicate in a secure manner. Each of the clients holds its own data that can be split into the training and evaluation subsets, and is responsible to train its own model. At the same time, the Server needs to have access to similar data to validate the aggregated model results. The model architecture is consistent across all participants, a necessary requirement for ensuring the compatibility of output dimensions and facilitating effective aggregation.

Another critical aspect of the system in the research setting is its flexibility in simulating and supporting different data partition types, including IID, NonIID, and Vertical partitions. In practice, this means that each participant must be able to adapt the dataset, based on the partition type PT defined in the Config Module.

- *Train* function represents a regular model training cycle on each client premises. The implementation can depend on a particular use-case, but in our case the training cycle is identical across all clients. That is, the parameters like learning rate, batch size, and number of epochs have identical values.
- *Evaluate* function measures the performance of the model on the testing dataset on each client. It uses common evaluation functions such as MSE and MAE of the locally trained model. These metrics, together with the model weight, are being sent to the server.
- *Aggregate* function encapsulates the process of combining the model weights. In our approach, regular and secure aggregation can be used. That is, FedAvg is used for regular aggregation and SecAgg+ for secure aggregation. The output of this process are aggregated model weights that contain features learned from each client.
- *Validate* function is analogous to Evaluate, except that it operates on a model based on aggregated weights. Additionally, it validates the combined model on the Servers' validation set.

Currently, the system is designed to execute a FL process with a single set of settings at a time. Consequently, to facilitate future evaluation and comparison, the model and its corresponding convergence history must be stored on the Server.

Chapter 4

Enhancing Privacy in Federated Learning: Method Assessment and Evaluation

This section discusses the application of Federated Learning for secure network monitoring by investigating its use in various data partitioning settings to train Deep Learning models across multiple data partitions, incorporating secure aggregation to enhance data privacy. We conduct experiments in two categories - FL with regular and secure secure aggregations - evaluating each across three data partition types: Non-Independent and Identically Distributed (NonIID), Independent and Identically Distributed (IID) and Vertical.

This setup allows to evaluate the impact of different data partitioning and aggregation methods on the final model's performance in the network monitoring setting. Moreover, we measure the computation time of our approach to assess the feasibility and general performance implications of FL across the partition types. This allows us to evaluate the impact of secure aggregation during the entire execution process.

4.1 Federated Learning with Regular Aggregation

The results of FL with regular aggregation evaluation on different data partitions is discussed. For horizontal partitions (IID and NonIID), the FL process involves three clients, whereas vertical partitions is limited to two participants.

The total FL process time for both partition types takes 193 seconds. Average time elapsed for each round is 18 seconds, except for the first and second, which take 24.5 and 21.4 seconds, respectively. The increased time of the first and second rounds is caused by the initialization and setup of the federated process, which requires exchanging the initial parameters and metadata with each client. Since both FL processes use the

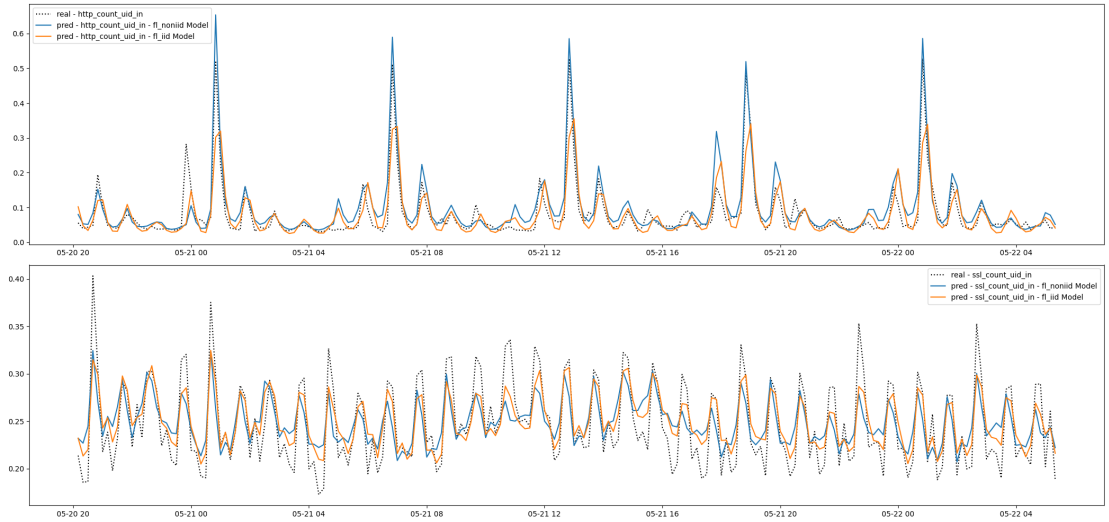


Figure 4.1: Predictions of the models trained on NonIID and IID data distributions. The horizontal axis represents the sample timestamps, and vertical axis - real and predicted activity level values for each protocol.

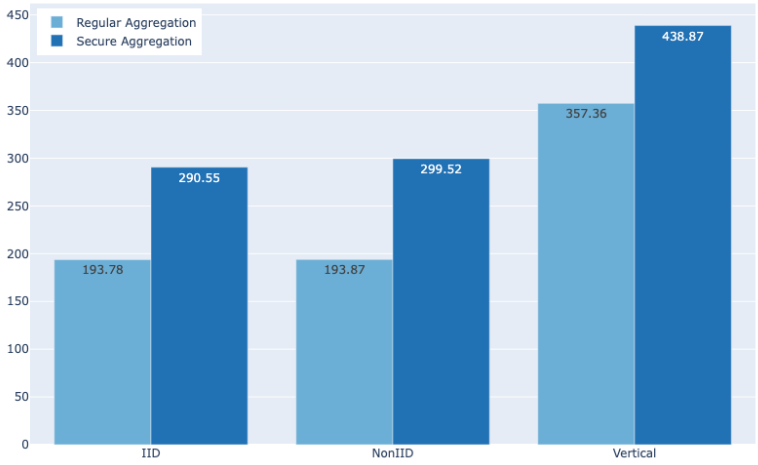


Figure 4.2: Total time in seconds elapsed for different FL settings

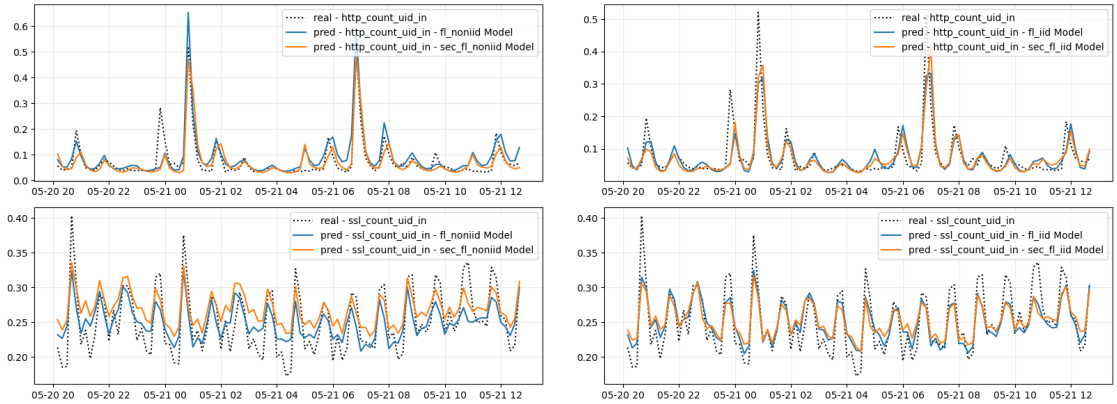
same setting and partition size, the equality of total and per-round elapsed time is expected.

4.2 Federated Learning with Secure Aggregation

This section details FL experiments using the SecAgg+ protocol to securely aggregate weights and preserve privacy. Similarly to the previous section, two Horizontal and

Vertical partitions are evaluated. The general settings for data partitioning, model, and federate learning remain unchanged from Section 4.1.

Horizontal Partitions The convergence of the model indicates that the secure aggregation reaches an acceptable MSE of 0.002324 for NonIID and 0.002941 for IID partitions, which is comparable to the regular approach. For secure NonIID, this represents the 1.13% increase from its regular counterpart, and for secure IID – 12.08% increase. Other metrics in the Tab. **??** have the same tendency to be 1% to 10% worse in the secure compared to regular. Similarly, per-round metrics are 1% - 10% percent higher across all rounds for secure NonIID in comparison to the regular NonIID model. The same applies to IID models, where an increase is observed between 2% and 12%. However, the model with secure aggregation converges more gradually than the regular models. The secure IID setting shows instability on the client models at the beginning of the FL process, although all client models were able to converge to the optimal MSE values at the end.



(a) Regular NonIID vs Secure NonIID

(b) Regular IID vs Secure IID

Figure 4.3: Predictions of the models trained on the horizontal distributions, with and without Secure Aggregation.

The prediction of regular and secure horizontal models is shown in Figures 4.3a and 4.3b. Comparison of actual and predicted values of the $http\_count\_uid\_in$ attribute using the cosine metric yields the 0.89 similarity for the secure NonIID model and the 0.90 for the regular NonIID model, which is 1% decrease in prediction accuracy. For the $ssl\_count\_uid\_in$ attribute, the predictions of the secure model capture the overall network load trend but struggle with data variability, leading to a 7.47% higher prediction value compared to the regular model, while both models achieve a cosine similarity of 0.97, confirming the closeness of their predictions. Experimental results indicate modest but consistent performance degradation when employing secure aggregation relative to standard federated learning.

The predictions of the secure IID model closely resemble those of the regular IID model for both features. Specifically, cosine similarity to actual values for $http\_count\_uid\_in$ attribute is 0.84 for the secure model and 0.86 for the regular model, which indicates a decrease of 2.32%, but confirms the usability of the secure model. For the $ssl\_count\_uid\_in$ attribute, both models perform similarly and have the 0.97 value of the cosine metric, which is considered a negligible impact of secure aggregation on the overall performance of the model.

The total time elapsed for FL training with secure aggregation is 290.55 seconds on NonIID partitions and 299.52 seconds on IID partitions, reflecting increases of 54.56% and 49.93%, respectively, compared to regular aggregation. The visual time comparison is shown in Figure 4.2. The time elapsed in each round on average is $50.25 - 55.24\%$ higher than in regular models. Similarly, the first and second rounds take longer due to the initialization and setup of the federated process.

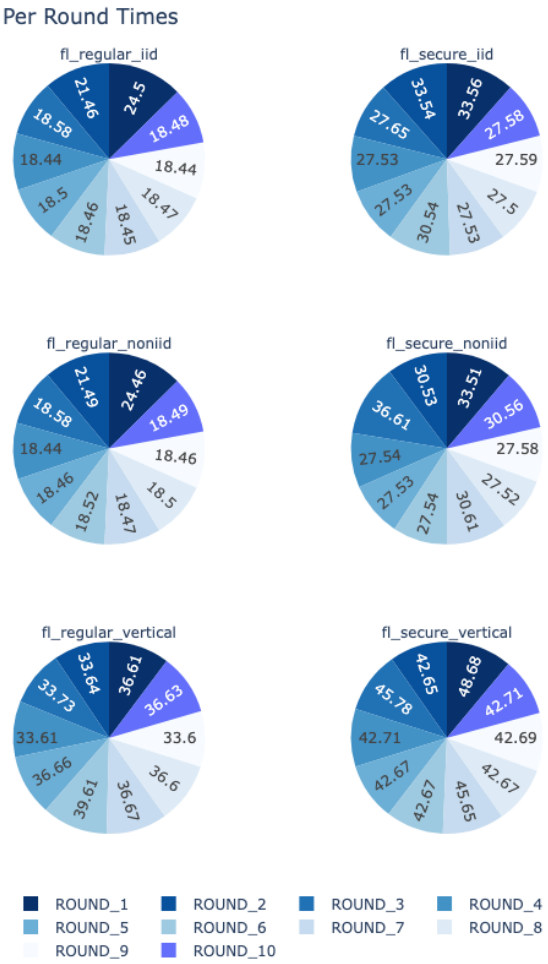


Figure 4.4: Time elapsed for each round for different FL settings

Vertical Partitions Model's convergence has more noticeable differences between regular and secure aggregation methods compared to horizontal partitions. Here, secure aggregation shows a higher loss in rounds 2, 3, 8, and the final round 10, but a lower loss in the remaining rounds, while maintaining a convergence gradient similar to the regular model. The average difference between secure and regular aggregation losses is in the range of $0.3 - 30.02\%$ with the final loss 37.53% higher in secure aggregation. The final MSE metrics is 0.000709 for $http\_count\_uid\_in$ and 0.000709 for $ssl\_count\_uid\_in$. Both clients follow a similar convergence pattern, reaching acceptable MSE values (0.022 and 0.019) in the third round and maintaining comparable results throughout the FL process.

The predictions, visualized in Figure 4.5, show the general similarity in model behavior between the regular and secure models. The secure model is capable of predicting the $http\_count\_uid\_in$ attribute with the same accuracy as the regular on, with a deviation up to 20% in certain sections. The cosine metric for the secure model is 0.97, while for the regular model is 0.98, which is only 1% percent decrease. However, the predictions for the $ssl\_count\_uid\_in$ attribute look less promising, where the secure model is unable to fit the minimum and maximum bounds of the initial sequence. Figure 4.5 clearly visualizes the difference, although the percentage difference is in the range $4.8 - 5.2\%$. The cosine metric for $ssl\_count\_uid\_in$ shows 0.99 for both secure and regular models, which confirms the models' ability to learn properly with secure aggregation.

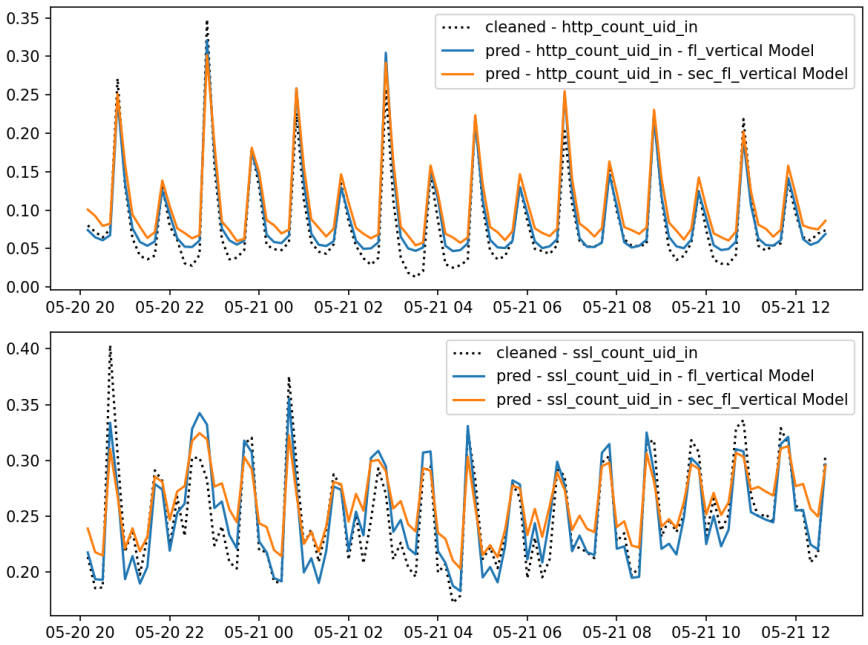


Figure 4.5: Predictions of the models on the vertical data distribution with and without Secure Aggregation.

In the context of time, the complete secure FL process takes 438.87 seconds, representing a 22.81% increase compared to the regular process. The visual comparison of the elapsed time is shown in Figure 4.2. The longest round took 48.69 seconds, and the other rounds averaged at 43.35 seconds per round. The round-wise times results represent an increase 21.65% compared to the regular vertical FL.

4.3 Summary

The proposed approach operates in the cybersecurity domain, where multiple clients collect network monitoring information and collaboratively train model for proactive network activity monitoring. Core building blocks are Federated Learning, privacy-preserving by design collaborative learning approach, and Secure Multi-Party Computation, which enhances the privacy by securing the aggregation process and prevents unwanted access to shared artifacts. System was evaluated in different data partitioning settings to simulate real-world usage, and in all settings, the system produced the model that significantly outperforms the one trained on a single client.

Achieved results demonstrate the feasibility and advantages of the FL-based system in learning from information without direct access while preserving the privacy of artifacts, such as raw data and model weights. The system demonstrates adaptability across diverse problem domains and data partitioning configurations, thereby improving its generalizability and practical applicability. This inherent flexibility is critical for real-world deployments, which frequently encounter heterogeneous data distributions and varying computational constraints, necessitating adaptable and scalable machine learning solutions.

Chapter 5

Conclusion

Sensitive data are an inseparable part of every modern-world operation, while machine learning allows learning from them. Accessing siloed data is problematic due to its nature, hindering the model improvements and making collaboration between several entities impossible. Through case studies and preliminary research, the thesis focused on Secure Multi-Party Computation, Federated Learning, and Differential Privacy as the most prominent and balanced approaches to securing data in machine learning.

Hybrid approach is based on Federated Learning, inherently avoiding a direct sharing of sensitive information among collaborating entities, while still enabling all participants to benefit from the collective dataset. Secure Multi-Party Computation is integrated for secure weight aggregation. These ensure the impossibility of data exposure during intermediate steps to participating parties or the coordinating server. Importantly, secure aggregation introduces only moderate computational overhead at an acceptable level, making it practical for real-world deployments when combined with optimization strategies. The proposed solution has proven its ability to successfully accommodate multiple data partitioning settings with a global model that exceeds single-client training in all metrics.

The works main contributions can be summarized as follows:

- Multiple strategies for preserving the privacy of artifacts in the collaborative machine learning process are explored, including Differential Privacy, Secure Multi-Party Computation for model training, and secure aggregation techniques.
- A secure federated learning approach for collaborative network monitoring is proposed, enabling joint model training without requiring full mutual trust while preserving the privacy of participants' data.
- Proposed approach proved its practical applicability in real-world condition by being able to accommodate different data partitions and introducing computation overhead at acceptable level.

Chapter 6

List of Publications

Journal papers

1. (A+, V3, ADC, JCR-Q2, SJR-Q1, IF2023=3.4, 1 citation)
   Oleksandr Lytvyn and Giang Nguyen. "Secure Federated Learning for Multi-Party Network Monitoring". In: *IEEE Access* 12 (2024). CC BY-NC-ND 4.0, pp. 163262–163284. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3486810
   Cited in:

   - Shubhi Shukla et al. "Federated learning with differential privacy for breast cancer diagnosis enabling secure data sharing and model integrity". In: *Scientific Reports* 15.1 (2025), p. 13061. DOI: 10.1038/s41598-025-95858-2

2. (A+, V3, ADC, JCR-Q1 decile, SJR-Q1, IF2024=13.9, 3 citations)
   Giang Nguyen, Judith Sáinz-Pardo Díaz, Amanda Calatrava, Lisana Berberi, Oleksandr Lytvyn, Valentin Kozlov, Viet Tran, Germán Moltó, and Álvaro López García. "Landscape of machine learning evolution: privacy-preserving federated learning frameworks and tools". In: *Artificial Intelligence Review* 58.2 (2025). Article 51, CC BY 4.0, p. 51. DOI: 10.1007/s10462-024-11036-2
   Cited in:

   - Jiqun Zhang et al. "A data trading scheme based on blockchain and game theory in federated learning". In: *Expert Systems with Applications* (2025), p. 127158. DOI: 10.1016/j.eswa.2025.127158
   - Catarina Silva, Joao Paulo Barraca, and Paulo Salvador. "Evaluating the Effectiveness of Differential Privacy Against Profiling". In: *Proceedings of the 2025 The International Conference on Consumer Technology (ICCT-Europe 2025), Algarve, Portugal*. 2025, pp. 28–30
   - M Swapna, Tanishqa Ravirala, and Nikhitha Reddy. "Diabetic Retinopathy Detection using Federated Learning and Vision Transformers". In: *International Journal of Interpreting Enigma Engineers (IJIEE)* 2.1 (2025), pp. 10–21

Conference papers

1. (A-, V2, AFC, 5 citations)
   Oleksandr Lytvyn and Giang Nguyen. "Secure Multi-Party Computation for Magnetic Resonance Imaging Classification". In: *Procedia Computer Science* 220 (2023). CC BY 4.0., 14th International Conference on Ambient Systems, Networks and Technologies (ANT'2023), pp. 24–31. ISSN: 1877-0509. DOI: 10.1016/j.procs.2023.03.006
   Cited in:

   - Sanjiv M Narayan, Nitin Kohli, and Megan M Martin. "Addressing contemporary threats in anonymised healthcare data using privacy engineering". In: *Nature npj Digital Medicine* 8.1 (2025), p. 145. DOI: 10.1038/s41746-025-01520-6
   - Ahmed Tamer Ahmed Hossam. "Enhancing Privacy in Big Data Analytics Through Encrypted Computational Techniques and Secure Multi-Party Computation Strategies". In: *Journal of Industrial IoT Technologies* 14.2 (2024), pp. 1–8. URL: %7Bhttps://scicadence.com/index.php/Industrial-IoT/article/view/37%7D
   - Laçi Hafsa and Sevrani Kozeta. "Preserving privacy in medical images while still enabling AI-driven research: A comprehensive review". In: *2024 13th Mediterranean Conference on Embedded Computing (MECO)*. IEEE. 2024, pp. 1–5. DOI: 10.1109/MECO62516.2024.10577795
   - Yicheng Gong, Wenlong Wu, and Linlin Song. "GAN-Based Privacy-Preserving Intelligent Medical Consultation Decision-Making". In: *Group Decision and Negotiation* (2024), pp. 1–28. DOI: 10.1007/s10726-024-09902-z
   - Thomas Buchsteiner et al. *webSPDZ: Versatile MPC on the Web*. Cryptology ePrint Archive, Paper 2025/487. 2025. URL: https://eprint.iacr.org/2025/487

2. (A, V2, AFC, CORE-B conference, 2 citations)
   Oleksandr Lytvyn and Giang Nguyen. "Efficiency and Security Trade-offs of Secure Multi-Party Computation for Machine Learning". In: *Procedia Computer Science* 225 (2023). CC BY 4.0., 27th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES'2023), pp. 655–664. ISSN: 1877-0509. DOI: 10.1016/j.procs.2023.10.051
   Cited in:

   - Yicheng Gong, Wenlong Wu, and Linlin Song. "GAN-Based Privacy-Preserving Intelligent Medical Consultation Decision-Making". In: *Group Decision and Negotiation* (2024), pp. 1–28. DOI: 10.1007/s10726-024-09902-z
   - Nzenwata Uchenna Jeremiah et al. "A Systematic Review of Privacy-preserving Techniques in Databases". In: *Asian Journal of Research in Computer Science*

18.7 (2025), pp. 38–48. DOI: 10.9734/ajrcos/2025/v18i7718

3. (A, V2, AFC, CORE-B conference, 1 citations)
Hlib Kokin, Oleksandr Lytvyn, and Giang Nguyen. "Optimal Differential Privacy for Deep Learning Model Training". In: *Procedia Computer Science* 246 (2024). CC BY 4.0., 28th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES'2024), pp. 2419–2428. ISSN: 1877-0509. DOI: 10.1016/j.procs.2024.09.483 Cited in:

   - Audris Arzovs et al. "Application of differential privacy to sensor data in water quality monitoring task". In: *Ecological Informatics* (2025), p. 103019. DOI: https://doi.org/10.1016/j.ecoinf.2025.103019

References

[1] EU GDPR. *General Data Protection Regulation*. Accessed on 03.04.2025, European Union. 2018. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[2] EU DSA Act. *European Union Digital Services Act*. Accessed on 03.04.2025, European Union. 2022. URL: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en.

[3] EU AI Act. *European Union Artificial Intelligence Act*. Accessed on 03.04.2025, European Union. 2024. URL: https://artificialintelligenceact.eu/the-act/.

[4] Chen Zhang et al. "A survey on federated learning". In: *Knowledge-Based Systems* 216 (2021), p. 106775. DOI: 10.1016/j.knosys.2021.106775.

[5] David Evans, Vladimir Kolesnikov, Mike Rosulek, et al. "A pragmatic introduction to secure multi-party computation". In: *Foundations and Trends® in Privacy and Security* 2.2-3 (2018), pp. 70–246. DOI: 10.1561/3300000019.

[6] Ahmed El Ouadrhiri and Ahmed Abdelhadi. "Differential privacy for deep and federated learning: A survey". In: *IEEE access* 10 (2022), pp. 22359–22380. DOI: 10.1109/ACCESS.2022.3151670.

[7] Giang Nguyen et al. "Landscape of machine learning evolution: privacy-preserving federated learning frameworks and tools". In: *Artificial Intelligence Review* 58.2 (2025). Article 51, CC BY 4.0, p. 51. DOI: 10.1007/s10462-024-11036-2.

[8] Qiang Yang et al. "Federated machine learning: Concept and applications". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 10.2 (2019), pp. 1–19. DOI: 10.1145/3298981.

[9] Cynthia Dwork, Aaron Roth, et al. "The algorithmic foundations of differential privacy". In: *Foundations and Trends in Theoretical Computer Science* 9.3-4 (2014), pp. 211–407. DOI: 10.1561/0400000042.

[10] Chuan Zhao et al. "Secure multi-party computation: theory, practice and applications". In: *Information Sciences* 476 (2019), pp. 357–372. DOI: 10.1016/j.ins.2018.10.024.

[11] Yehuda Lindell. "Secure multiparty computation". In: *Communications of the ACM* 64.1 (2020), pp. 86–96. DOI: 10.1145/3387108.

[12] Ronald Cramer, Ivan Damgård, and Jesper B Nielsen. "Multiparty computation from threshold homomorphic encryption". In: *Advances in Cryptology 2001: Inter-*

national Conference on the Theory and Application of Cryptographic Techniques Innsbruck, Austria, May 6–10, 2001 Proceedings 20. Springer. 2001, pp. 280–300. DOI: 10.1007/3-540-44987-6_18.

[13] Arpita Patra and Akshayaram Srinivasan. "Three-round secure multiparty computation from black-box two-round oblivious transfer". In: *Advances in Cryptology–CRYPTO 2021: 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16–20, 2021, Proceedings, Part II 41*. Springer. 2021, pp. 185–213. DOI: 10.1007/978-3-030-84245-1_7.

[14] Yuval Ishai et al. "Zero-knowledge proofs from secure multiparty computation". In: *SIAM Journal on Computing* 39.3 (2009), pp. 1121–1152. DOI: 10.1137/080725398.

[15] Elette Boyle, Niv Gilboa, and Yuval Ishai. "Function secret sharing". In: *Annual international conference on the theory and applications of cryptographic techniques*. Springer. 2015, pp. 337–367. DOI: 10.1007/978-3-662-46803-6_12.

[16] Payman Mohassel and Peter Rindal. "ABY3: A mixed protocol framework for machine learning". In: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. 2018, pp. 35–52. DOI: 10.1145/3243734.3243760.

[17] Sameer Wagh et al. "Falcon: Honest-majority maliciously secure framework for private deep learning". In: *arXiv preprint arXiv:2004.02229* (2020). DOI: 10.48550/arXiv.2004.02229.

[18] Luke A Bauer and Vincent Bindschaedler. "Towards realistic membership inferences: The case of survey data". In: *Annual Computer Security Applications Conference*. 2020, pp. 116–128. DOI: 10.1145/3427228.3427282.

[19] Xuejun Zhao et al. "Exploiting explanations for model inversion attacks". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 682–692. DOI: 10.1109/ICCV48922.2021.00072.

[20] Mohammad Malekzadeh, Anastasia Borovykh, and Deniz Gündüz. "Honest-but-curious nets: Sensitive attributes of private inputs can be secretly coded into the classifiers' outputs". In: *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. 2021, pp. 825–844. DOI: 10.1145/3460120.3484533.

[21] Xueluan Gong et al. "InverseNet: Augmenting Model Extraction Attacks with Training Data Inversion." In: *IJCAI*. 2021, pp. 2439–2447. URL: https://www.ijcai.org/proceedings/2021/0336.pdf.

[22] Keith Bonawitz et al. "Practical secure aggregation for privacy-preserving machine learning". In: *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 2017, pp. 1175–1191. DOI: 10.1145/3133956.3133982.

[23] Giang Nguyen et al. "Network security AIOps for online stream data monitoring". In: *Neural Computing and Applications* (2024), pp. 1–25. DOI: 10.1007/s00521-024-09863-z.

[24] Giang Nguyen et al. "Deep learning for proactive network monitoring and security protection". In: *IEEE Access* 8 (2020), pp. 1–21. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2968718.

[25] Oleksandr Lytvyn and Giang Nguyen. "Secure Federated Learning for Multi-Party Network Monitoring". In: *IEEE Access* 12 (2024). CC BY-NC-ND 4.0, pp. 163262–163284. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3486810.

[26] Shubhi Shukla et al. "Federated learning with differential privacy for breast cancer diagnosis enabling secure data sharing and model integrity". In: *Scientific Reports* 15.1 (2025), p. 13061. DOI: 10.1038/s41598-025-95858-2.

[27] Jiqun Zhang et al. "A data trading scheme based on blockchain and game theory in federated learning". In: *Expert Systems with Applications* (2025), p. 127158. DOI: 10.1016/j.eswa.2025.127158.

[28] Catarina Silva, Joao Paulo Barraca, and Paulo Salvador. "Evaluating the Effectiveness of Differential Privacy Against Profiling". In: *Proceedings of the 2025 The International Conference on Consumer Technology (ICCT-Europe 2025), Algarve, Portugal*. 2025, pp. 28–30.

[29] M Swapna, Tanishqa Ravirala, and Nikhitha Reddy. "Diabetic Retinopathy Detection using Federated Learning and Vision Transformers". In: *International Journal of Interpreting Enigma Engineers (IJIEE)* 2.1 (2025), pp. 10–21.

[30] Oleksandr Lytvyn and Giang Nguyen. "Secure Multi-Party Computation for Magnetic Resonance Imaging Classification". In: *Procedia Computer Science* 220 (2023). CC BY 4.0., 14th International Conference on Ambient Systems, Networks and Technologies (ANT'2023), pp. 24–31. ISSN: 1877-0509. DOI: 10.1016/j.procs.2023.03.006.

[31] Sanjiv M Narayan, Nitin Kohli, and Megan M Martin. "Addressing contemporary threats in anonymised healthcare data using privacy engineering". In: *Nature npj Digital Medicine* 8.1 (2025), p. 145. DOI: 10.1038/s41746-025-01520-6.

[32] Ahmed Tamer Ahmed Hossam. "Enhancing Privacy in Big Data Analytics Through Encrypted Computational Techniques and Secure Multi-Party Computation Strategies". In: *Journal of Industrial IoT Technologies* 14.2 (2024), pp. 1–8. URL: %7Bhttps://scicadence.com/index.php/Industrial-IoT/article/view/37%7D.

[33] Laçi Hafsa and Sevrani Kozeta. "Preserving privacy in medical images while still enabling AI-driven research: A comprehensive review". In: *2024 13th Mediterranean Conference on Embedded Computing (MECO)*. IEEE. 2024, pp. 1–5. DOI: 10.1109/MECO62516.2024.10577795.

[34] Yicheng Gong, Wenlong Wu, and Linlin Song. "GAN-Based Privacy-Preserving Intelligent Medical Consultation Decision-Making". In: *Group Decision and Negotiation* (2024), pp. 1–28. DOI: 10.1007/s10726-024-09902-z.

[35] Thomas Buchsteiner et al. *webSPDZ: Versatile MPC on the Web*. Cryptology ePrint Archive, Paper 2025/487. 2025. URL: https://eprint.iacr.org/2025/487.

[36] Oleksandr Lytvyn and Giang Nguyen. "Efficiency and Security Trade-offs of Secure Multi-Party Computation for Machine Learning". In: *Procedia Computer Science* 225 (2023). CC BY 4.0., 27th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES'2023), pp. 655–664. ISSN: 1877-0509. DOI: 10.1016/j.procs.2023.10.051.

[37] Nzenwata Uchenna Jeremiah et al. “A Systematic Review of Privacy-preserving Techniques in Databases”. In: *Asian Journal of Research in Computer Science* 18.7 (2025), pp. 38–48. DOI: 10.9734/ajrcos/2025/v18i7718.

[38] Hlib Kokin, Oleksandr Lytvyn, and Giang Nguyen. “Optimal Differential Privacy for Deep Learning Model Training”. In: *Procedia Computer Science* 246 (2024). CC BY 4.0., 28th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES’2024), pp. 2419–2428. ISSN: 1877-0509. DOI: 10.1016/j.procs.2024.09.483.

[39] Audris Arzovs et al. “Application of differential privacy to sensor data in water quality monitoring task”. In: *Ecological Informatics* (2025), p. 103019. DOI: https://doi.org/10.1016/j.ecoinf.2025.103019.