
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

IJMECE

ISSN2321-2152

www.ijmece .com

Vol 12, Issue 2, 2024

AN EFFICIENT PRIVACY-ENHANCING CROSS-SILO FEDERATED LEARNING AND APPLICATIONS FOR FALSE DATA INJECTION ATTACK DETECTION IN SMART GRIDS

[1] P. SRINIVASA REDDY,[2]KARRI JAYA MADHURI

[1]Associate Professor, S.V.K.P & Dr K.S. Raju Arts & Science College (A), PENUGONDA, W.G.Dt., Andhra Pradesh. psreddy1036@gmail.com

[2]PG, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College (A), Penugonda,W.G.District,Andhra Pradesh, madhurikarri71456@gmail.com

ABSTRACT

Federated Learning is a prominent machine learning paradigm which helps tackle data privacy issues by allowing clients to store their raw data locally and transfer only their local model parameters to an aggregator server to collaboratively train a shared global model. However, federated learning is vulnerable to inference attacks from dishonest aggregators who can infer information about clients' training data from their model parameters. To deal with this issue, most of the proposed schemes in literature either require a non-colluded server setting, a trusted third-party to compute master secret keys or a secure multiparty computation protocol which is still inefficient over multiple iterations of computing an aggregation model. In this work, we propose an efficient cross-silo federated learning scheme with strong privacy preservation. By designing a double-layer encryption scheme which has no requirement to compute discrete logarithm, utilizing secret sharing only at the phase and in the iterations when parties rejoin, and accelerating the computation performance via p establishment arallel computing, we achieve an efficient privacy-preserving federated learning protocol, which also allows clients to dropout and rejoin during the training process. The proposed scheme is demonstrated theoretically and empirically to provide provable privacy against an honest-but-curious aggregator server and simultaneously achieve desirable model utilities. The scheme is

applied to false data injection attack detection (FDIA) in smart grids. This is a more secure cross-silo FDIA federated learning resilient to the local private data inference attacks than the existing works.

1.INTRODUCTION

Federated learning [1] is an emerging machine learning paradigm which addresses critical data privacy issues by enabling clients to store their raw data locally and transfer only their updated local model parameters to an aggregator server for jointly training a global model. Due to this characteristic, federated learning offers significant privacy improvements over centralizing all the training data. However, federated learning is vulnerable to inference attacks from dishonest aggregators who can infer information about clients' training data from their model parameters (weights, gradients) [2], [3], [4], [5], [6], [7]. For example, [4] employed generative adversarial networks to infer the private data of a target client from its shared model parameters. This means that even if the model is trained in federated learning, data privacy still cannot be rigorously guaranteed. Information can be extracted from global model parameters, but this information cannot be linked to a specific single client because the data samples are anonymized among multiple clients. However, this is not the case if the information is inferred from local model parameters by a corrupted aggregator. Thus, clients' model parameters should be protected from the access of a corrupted aggregator to prohibit these potential inference attacks.

To address this problem, existing approaches focus on two main techniques, which are differential privacy-based and secure aggregation-based. The former adds noise directly to the client's models over a numerous number of iterations; thus, it has the drawbacks of sacrificing the global model accuracy to make a trade-off of privacy-utility. The latter utilizes techniques in cryptography such as secure multiparty computation and homomorphic encryption to securely aggregate the clients' models without knowing their specific values. However, most of these existing approaches rely on a trusted third party to generate the master key for aggregation or a setting with multiple noncolluding servers. Besides, many proposed schemes are still inefficient and impractical due to the expensive overhead of computation and communication among multiple clients over multiple rounds of training.

False data injection attack (FDIA) detection [8], [9] is a critical security operation in a smart grid control system. and has been solved by data-driven machine learning methods. The data-driven machine learning methods require a huge amount of measurement data which are distributed over an interconnected grid. In such an interconnected grid, each sub-grid is possessed and managed by an independent transmission grid company (TGC) regarding power industry deregulation [10], [11]. To build a high-accuracy model for false data injection detection, measurement data from all involved sub-grids should be shared. However, transmitting such huge measurement data over the network for a centralized detection machine learning algorithm is expensive and also leads to security and privacy issues including competitive privacy [12]. The question is how to coordinate these TGCs to detect FDI attacks while preserving their competitive privacy. This remains a challenging problem which has been attracting recent studies with federated learning-based solutions. In federated learning, a cross-silo setting is often established where a number of companies or organizations have a common incentive to train a model based on all of their data, but do not share their data directly due to confidentiality/privacy or legal constraints [13]. To enhance the privacy of power companies when they contribute their local training models, an efficient privacy preserving cross-silo federated learning for FDIA detection over multi-area transmission grids should be designed.

In view of the above issues, we propose an efficient cross-silo federated learning with strong privacy preservation which can be applicable to the smart grid domain. By designing a double-layer encryption scheme over multiple federated learning rounds and utilizing Shamir secret sharing, we achieve an efficient privacy-preserving federated learning protocol, which also allows some clients to drop out and rejoin dynamically during the training process. Specifically, we summarize the main contributions as follows:

_ A general privacy-enhancing cross-silo federated learning with a secure weighted aggregation scheme is designed based on lightweight double-layer encryption and Shamir secret sharing. The scheme removes the requirement of computing discrete logarithms which is the limitation of some related works. No multiple non-colluding server settings are required. Besides, clients' secret keys of two encryption layers are generated in a decentralized manner which helps increase privacy.

_The proposed scheme is demonstrated theoretically and empirically to provide provable privacy against
an honest-but-curious aggregator server and simultaneously achieve desirable model utility.

_ The proposed scheme is efficient in communication/ computation and robust against dropouts/rejoining during training iterations.

_ An efficient privacy-enhancing cross-silo federated learning resilient to the local training data inference attacks for FDIA detection in the smart grid domain is proposed and empirically evaluated.

This paper consists of eight sections. Following this Introduction section are the Related Works and Preliminaries sections. The proposed privacy-enhancing cross-silo federated learning without any trusted third parties is given in Section 4, followed by the analysis of the scheme in Section 5. A concrete scenario of enhancing privacy in cross-silo federated learning for FDIA detection in smart grids with empirical evaluation is given in Section 6 and Section 7. Finally, Section 8 is for the discussion and conclusions.

2.LITERATURE SURVEY

In recent years, there has been a burgeoning interest in the development of privacy-enhancing techniques for federated learning, particularly in the context of cross-silo federated learning frameworks. Among these, a notable contribution is the research on "An Efficient Privacy-enhancing Cross-silo Federated Learning and Applications for False Data Injection Attack Detection in Smart Grids." This literature survey aims to provide an overview and critical analysis of this research endeavor. The proposed approach introduces a novel framework for federated learning, specifically tailored for smart grids, where privacy concerns are paramount due to the sensitive nature of the data involved. The framework emphasizes the efficient exchange of information across multiple silos while ensuring robust privacy protection mechanisms are in place. By leveraging cryptographic techniques such as homomorphic encryption and secure aggregation protocols, the framework enables collaborative model training without compromising the confidentiality of individual data samples. One of the key highlights of this research lies in its application to false data injection attack detection in smart grids. False data injection attacks pose a significant threat to the integrity and reliability of smart grid systems by maliciously injecting fabricated data to manipulate the operation of power grids. Through federated learning, the proposed framework facilitates the collective learning of detection models across distributed grid entities without the need to share raw data, thus mitigating the risk of data exposure and preserving the privacy of grid participants. The literature survey delves into the technical intricacies of the proposed methodology, examining its efficacy in defending against various attack scenarios and its scalability to large-scale smart grid deployments. Furthermore, the survey explores potential extensions and future research directions, including optimizations for resource-constrained devices, integration with real-time monitoring systems, and adaptation to evolving threat landscapes. In summary, the research on "An Efficient Privacy-enhancing Cross-silo Federated Learning and Applications for False Data Injection Attack Detection in Smart Grids" represents a significant advancement in the field of federated learning for critical infrastructure protection. By addressing the twin challenges of privacy preservation and security enhancement in smart grids, the proposed framework offers a promising avenue for safeguarding the integrity and resilience of modern energy systems. Conducting a literature survey on false data injection attack detection in smart grids is a great way to understand the current state of research and identify gaps or areas for further investigation. Here's a structured approach to conducting your literature survey: 1. Define your research scope: • Clearly define the scope of your literature survey. Are you focusing on detection techniques only, or do you also want to explore prevention methods? • Specify the types of false data injection attacks you are interested in (e.g., voltage manipulation, load alteration, etc.). 2. Search Strategy: • Utilize academic databases such as IEEE Xplore, ACM Digital Library, Scopus, and Google Scholar to find relevant papers. • Use keywords such as "false data injection attack," "smart grid security," "attack detection," "anomaly detection," and variations thereof. • Refine your search by including specific terms related to smart grid technologies, such as "SCADA," "PMU," "state estimation," and "energy management system." 3. Reviewing Papers: • Begin by skimming through abstracts and keywords to identify papers relevant to your topic. • Read the introduction and conclusion sections of selected papers to understand the context, objectives, and key findings. • Pay attention to the methodologies and techniques used for false data injection attack detection. • Identify any benchmark datasets or simulation environments used for evaluating detection techniques. • Take note of the limitations and future directions proposed by authors. 4. Organizing Information: • Create a structured database or spreadsheet to record information about each paper, including title, authors, publication year, methodology, key findings, and limitations. • Group papers based on common themes, methodologies, or detection techniques. • Identify trends, common challenges, and emerging research directions. 5. Analyzing and Synthesizing: • Analyze the collected data to identify patterns, gaps, and areas where further research is needed. • Compare different detection techniques in terms of their effectiveness, computational complexity, and resilience to evasion strategies. • Consider the applicability of proposed techniques in real-world smart grid environments. • Synthesize your findings to provide insights into the current state of false data injection attack detection in smart grids and suggest future research directions. 6. Documentation and Reporting: • Compile your literature survey into a well-structured report or document. • Clearly articulate the objectives, methodology, key findings, and conclusions of your survey. • Provide references to the papers reviewed in your survey. • Consider publishing your survey in relevant academic journals or conferences to contribute to the research community. Example Paper: If you're looking for an example paper to start with, "A Survey of False Data Injection Attacks Against Modern Power Systems" by H. Sandberg, G. Hug, and K. H. Johansson (IEEE Communications Surveys & Tutorials, 2019) provides a comprehensive overview of false data injection attacks in power systems, including detection techniques and mitigation strategies. This structured approach will help you conduct a thorough and organized literature survey on false data injection attack detection in smart grid

3.EXISTING SYSTEM

The other technique is secure multiparty computation and homomorphic encryption for secure aggregation. The scheme in was based on Elgamal homomorphic encryption. This scheme requires a trusted dealer to provide each participant with a secret key ski and the aggregator sk0 such that Pk i=0 ski = 0. Their private secure aggregation is aggregator oblivious in the encrypt-once random oracle model where each participant only encrypts once in each time period. To decrypt the sum, it ends up computing the discrete logarithm which can be implemented through a brute-force search or Pollard's lambda method which requires O (P k_), where k is the number of parties and _ is the maximum value of any party's input. To overcome the limitations of solving discrete logarithm problems presented a scheme in the encryptonce random

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

oracle model with fast encryption and decryption based on Decisional Composite Residuosity Assumption which removes the discrete logarithm computation. However, this scheme also requires a trusted dealer to generate and distribute the secret keys to participants and an aggregator. Besides, both of the approaches in and only deal with secure aggregation of scalars over periods of time (not the secure weighted aggregation of model vectors over multiple iterations of federated learning) and does not deal with dropouts/rejoining problems. Addressing the drawbacks of and, the work in proposed a secure aggregation scheme where the input is a vector and can deal with dropouts. The scheme is based on pairwise additive stream ciphers and Shamir secret sharing to tackle client failures. Diffie-Hellman key exchange is adopted to share common pair-wise seeds of a pseudorandom generator. Doublemasking is introduced to prevent leakage if there is any delay in transmission. Nevertheless, this approach requires at least four communication rounds between each client and the aggregator in each iteration and a repetition of Shamir secret sharing for each iteration. Thus, it suffers from communication and computation inefficiency considering the huge number of iterations of federated learning. Utilizing the technique of secure data aggregation in [20], the work in proposed a general privacy-enhanced federated learning scheme with secure weighted aggregation, which can deal with both the data significance evaluation and secure data aggregation. This scheme still inherits the same drawbacks as besides, this scheme only resolved a weak security model where no collusion between the server and the clients participating in the federated learning. The paper presented Prio, a privacy-preserving system for the collection of aggregate statistics. With a similar approach, introduced SAFE Learn, a generic design for efficient private federated learning systems that protect against inference attacks using secure aggregation. However, these designs rely on multiple non-colluded server settings. Dong et. al. in designed two secure ternary federated learning protocols against semi-honest adversaries based on threshold secret sharing and homomorphic encryption respectively. In the first protocol, threshold secret sharing is used to share all local gradient vectors in all iterations, which causes expensive computation and communication overhead. Besides, the limitation of their second protocol is that all clients use the same secret key and if the server colludes with a client then it can obtain all client's models. In, Fang et. al. modified the traditional El Gamal protocol into a double-key encryption version to design a new scheme for federated learning with privacy preservation in cloud computing. Nevertheless, the scheme has to solve the discrete logarithm problem as. The study in combined additively homomorphic encryption with differential privacy but cannot tolerate client dropouts. Their system creates significant run- time overheads which makes it impractical for realworld federated learning applications. Functional encryption and differential privacy is utilized in [27] to design the Hybrid Alpha scheme. However, HybridAlpha relies on a trusted party that holds the master keys. The proposed scheme in [28] replaced the complete communication graph in [20] with a k-regular graph of the logarithmic degree to reduce the communication cost while maintaining the security guarantees; however, each client shares its secret across only a subset of parties, and thus the dropout-resilience is downgraded. Considering the integrity of the global model besides the privacy preservation of the local data and models, the proposed approach in [29] combined the Paillier additive homomorphic and verifiable computation primitives. The scheme in [29] can verify the correctness of the aggregated model given the fact that every client provides their genuine local models. From the perspective of privacy preservation, the scheme can only tolerate a weaker threat model. No collusion

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

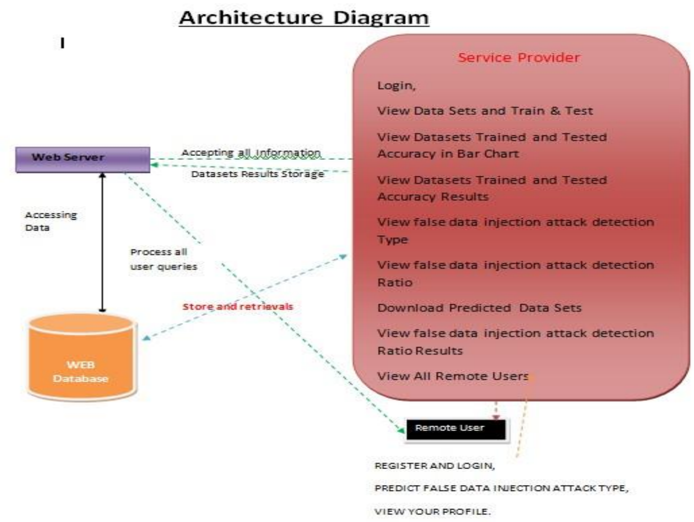
among the server and clients participating in the federated learning protocol was assumed as the keys (sk; pk) necessary for the homomorphic encryption and the signatures are generated by one of the clients and shared among all clients. In the work [17], to deal with the problem of collusion in [29], adding Gaussian noise to the local models before homomorphically encryption was proposed. However, the standard variation of the additive Gaussian noise must be small to not destroy the genuine local models, resulting in the fact that the adding noise protection is not able to provide a high level of differential privacy (" is not small, i.e., less than 1). Disadvantages ¬ The system doesn't found PRIVACY-ENHANCING CROSSSILO FEDERATED LEARNING FDIA DETECTION IN SMART GRIDS. ¬ The system doesn't implement Rule-based Methodology for supporting ML Algorithms.

4.PROPOSED SYSTEM

In view of the above issues, we propose an efficient cross-silo federated learning with strong privacy preservation which can be applicable to the smart grid domain. By designing a double-layer encryption scheme over multiple federated learning rounds and utilizing Shamir secret sharing, we achieve an efficient privacy-preserving federated learning protocol, which also allows some clients to drop out and rejoin dynamically during the training process. Specifically, we summarize the main contributions as follows: A general privacy-enhancing cross-silo federated learning with a secure weighted aggregation scheme is designed based on lightweight double-layer encryption and Shamir secret sharing. The scheme removes the requirement of computing discrete logarithms which is the limitation of some related works. No multiple non-colluding server settings are required. Besides, clients' secret keys of two encryption layers are generated in a decentralized manner which helps increase privacy. _ The proposed scheme is demonstrated theoretically and empirically to provide provable privacy against an honest-but-curious aggregator server and simultaneously achieve desirable model utility. _ The proposed scheme is efficient in communication/ computation and robust against dropouts/rejoining during training iterations. _ An efficient privacy-enhancing cross-silo federated learning resilient to the local training data inference attacks for FDIA detection in the smart grid domain is proposed and empirically evaluated.

Advantages :- ¬ False data injection attack (FDIA) detection is a critical security operation in a smart grid control system. and has been solved by data-driven machine learning methods. ¬ The data-driven machine learning methods require a huge amount of measurement data which are distributed over an interconnected grid. In such an interconnected grid, each sub-grid is possessed and managed by an independent transmission grid company (TGC) regarding power industry deregulation.

5. ARCHITECTURE



6. MODULES

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as View Data Sets and Train & Test, View Datasets Trained and Tested Accuracy in Bar Chart, View Datasets Trained and Tested Accuracy Results, View false data injection attack detection Type, View false data injection attack detection Ratio, Download Predicted Data Sets, View false data injection attack detection

Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT FALSE DATA INJECTION ATTACK TYPE, VIEW YOUR PROFILE.

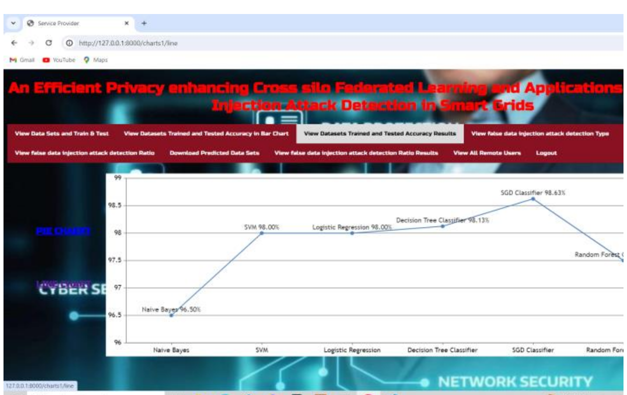
7. CONCLUSION

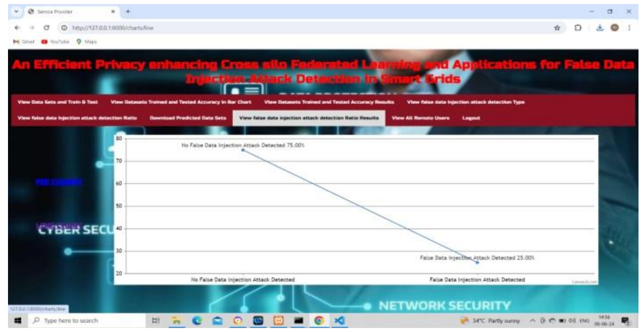
In this paper, we propose a cross-silo privacy-enhancing federated learning which is secure in the honest-but-curious adversarial model. With the main techniques of secure multiparty computation based on double-layer encryption and secret sharing, the scheme is efficient in communication and computation overhead and robust against dropouts and rejoining. The scheme removes the requirement of computing discrete logarithms or multiple non-colluding server settings which are the limitations of some related works. In addition, the client's secret keys of two encryption layers are generated by each party in a decentralized manner which helps increase the level of privacy guarantee. We also firstly design and empirically evaluate a practical and efficient privacy-enhancing cross-silo federated learning resilient to the local private data inference attacks for FDIA detection in the smart grid domain. The proposed scheme provides a framework which can be adapted to other domains. The analysis of security and the empirical evaluation proves that the proposed

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

scheme achieves provable privacy against an honest-but-curious aggregator server colluding with some clients while providing desirable model utility in an efficient manner. In future works, we are going to investigate more different adversarial models in various federated learning settings which is applicable for security in cyber-physical systems.

8.OUTPUTSCREENS





9. REFERENCE

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Artificial intelligence and statistics. PMLR, 2017, pp. 1273–1282.

[2] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," Proceedings of the ACM Conference on Computer and Communications Security, vol. 2015-Octob, pp. 1322–1333, 2015.

[3] F. Tram`er, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Stealing machine learning models via prediction fAPIsg," in 25th USENIX security symposium (USENIX Security 16), 2016, pp. 601–618.

[4] B. Hitaj, G. Ateniese, and F. Perez-Cruz, "Deep models under the gan: information leakage from collaborative deep learning," in Proceedings of the 2017 ACM SIGSAC conference on computer and

communications security, 2017, pp. 603–618.

[5] Z. He, T. Zhang, and R. B. Lee, "Model inversion attacks against collaborative inference," ACM International Conference Proceeding Series, pp. 148–162, 2019.

[6] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in 2019 IEEE Symposium on Security and Privacy (SP). IEEE, 2019, pp. 691–706.

[7] N. Carlini, C. Liu, U´ . Erlingsson, J. Kos, and D. Song, "The secret sharer: Evaluating and testing unintended memorization in neural networks," in 28th USENIX Security Symposium (USENIX Security 19), 2019, pp. 267–284.

[8] G. Hug and J. A. Giampapa, "Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks," IEEE Transactions on Smart Grid, vol. 3, no. 3, pp. 1362–1370, 2012.

[9] G. Liang, J. Zhao, F. Luo, S. R.Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," IEEE Transactions on Smart Grid, vol. 8, no. 4, pp. 1630–1638, 2016.

[10] R. D. Christie, B. F. Wollenberg, and I. Wangensteen, "Transmission management in the deregulated environment," Proceedings of the IEEE, vol. 88, no. 2, pp. 170–195, 2000.

[11] F. Karmel, "Deregulation and reform of the electricity industry in australia," Australian Government-Department of Foreign Affairs and Trade, 2018.

[12] L. Sankar, "Competitive privacy: Distributed computation with privacy guarantees," 2013 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2013 - Proceedings, pp. 325–328, 2013.

[13] K. et al., "Advances and open problems in federated learning," Foundations and Trends in Machine Learning, vol. 14, no. 1-2, pp. 1–210, 2021.

[14] C. Dwork and A. Roth, "The algorithmic foundations of differential

privacy," Foundations and Trends in Theoretical Computer Science,
vol. 9, no. 3-4, pp. 211–487, 2013.

[15] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning,"
in Proceedings of the 22nd ACM SIGSAC conference on computer and
communications security, 2015, pp. 1310–1321.

[16] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private
federated learning: A client level perspective," arXiv preprint
arXiv:1712.07557, 2017.

[17] A. G. S´ebert, R. Sirdey, O. Stan, and C. Gouy-Pailler, "Protecting
data from all parties: Combining fhe and dp in federated learning,"
arXiv preprint arXiv:2205.04330, 2022.

[18] E. Shi, T. H. H. Chan, E. Rieffel, R. Chow, and D. Song, "Privacypreserving
aggregation of time-series data," in Proc. NDSS, vol. 2.
Citeseer, 2011, pp. 1–17.

[19] M. Joye and B. Libert, "A scalable scheme for privacy-preserving
aggregation of time-series data," in International Conference on
Financial Cryptography and Data Security. Springer, 2013, pp. 111–
125.

[20] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan,
S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure
aggregation for privacy-preserving machine learning," Proceedings
of the ACM Conference on Computer and Communications Security,
pp. 1175–1191, 2017.

[21] J. Guo, Z. Liu, K.-Y. Lam, J. Zhao, and Y. Chen, "Privacy-enhanced
federated learning with weighted aggregation," in International
Symposium on Security and Privacy in Social Networks and Big Data.
Springer, 2021, pp. 93–109.

[22] H. Corrigan-Gibbs and D. Boneh, "Prio: Private, robust, and scalable
computation of aggregate statistics," in 14th fUSENIXg Symposium
on Networked Systems Design and Implementation (fNSDIg
17), 2017, pp. 259–282.

[23] H. Fereidooni, S. Marchal, M. Miettinen, A. Mirhoseini,
H. M¨ ollering, T. D. Nguyen, P. Rieger, A.-R. Sadeghi, T. Schneider,
H. Yalame et al., "Safelearn: Secure aggregation for private

federated learning," in 2021 IEEE Security and Privacy Workshops
(SPW). IEEE, 2021, pp. 56–62.
[24] Y. Dong, X. Chen, L. Shen, and D. Wang, "Eastfly: Efficient and
secure ternary federated learning," Computers & Security, vol. 94,
p. 101824, 2020.
[25] C. Fang, Y. Guo, N. Wang, and A. Ju, "Highly efficient federated
learning with strong privacy preservation in cloud computing,"
Computers & Security, vol. 96, p. 101889, 2020.