SCIENCE CHINA
Information Sciences

• LETTER •

April 2024, Vol. 67, Iss. 4, 149302:1–149302:2
https://doi.org/10.1007/s11432-023-3946-x

FUSE: a federated learning and U-shape split learning-based electricity theft detection framework

Xuan LI[1], Naiyu WANG[1], Liehuang ZHU[2], Shuai YUAN[3] & Zhitao GUAN[1*]

[1]*School of Control and Computer Engineering, North China Electric Power University, Beijing* 102206*, China;*
[2]*School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing* 100081*, China;*
[3]*Department of Finance, Operations, and Information Systems* (*FOIS*)*, Brock University, St.Catharines* L2S 3A1*, Canada*

Received 5 June 2023/Revised 16 November 2023/Accepted 26 January 2024/Published online 21 March 2024

In recent years, the data-driven electricity theft detection methods integrated with edge cloud computing [1, 2] have not only demonstrated superior detection accuracy but also improved efficiency, making them viable alternatives to indoor inspections. Energy service providers (ESPs) typically manage regions by dividing them into various transformer districts (TDs). The detection of electricity theft in a particular region is performed by the associated TD, which in turn determines whether a consumer is engaged in the theft behavior or not [3]. However, ensuring data privacy is a major challenge during the model training process on a cloud server (CS). This is because the fine-grained consumption data can potentially reveal undisclosed preference information of consumers, resulting in a reluctance from both consumers and ESPs to share their data.

As a distributed paradigm, federated learning is considered a promising solution that enables multiple data owners to jointly train a global model without exposing their raw data [4]. In a vanilla federated learning theft detection framework, TDs in different areas collect consumption data and train the detection model with the assistance of an aggregator. However, deploying a vanilla federated learning-based framework in the smart grid as a distributed alternative to train a theft detection model still presents certain challenges. Firstly, the privacy of consumption data needs to be preserved throughout the federated learning process. Meanwhile, the increasing scale of the model incurs a significant computational requirement on resource-constrained devices. Besides, the straggler issue caused by discrepancies in the hardware resources and communication link quality may have a negative impact on model performance. Therefore, we address the aforementioned problems by proposing a federated learning and U-shape split learning-based electricity theft detection framework (FUSE). The preliminary about federated learning and split learning is detailed in Appendix A.

Proposed framework. In our framework, we consider M TDs ($\mathrm{TD} = \{\mathrm{TD}_i | i \in [1, M]\}$) in different ESPs are involved in the federated learning process to train a theft detection model jointly. For each TD, there are N smart meters (SMs) SM_i^j ($j \in [1, N]$), which maintains the daily consumption data x_i^j of a given consumer. y_i^j represents the consumer label of x_i^j, denoting whether or not the theft behavior exists on SM_i^j. Thus, it is assigned and managed by TD_i. The detailed problem formulation and threat model are provided in Appendix B.

As shown in Figure 1, four kinds of entities participate in the FUSE framework: SMs, TDs, CSs, and an aggregator. The workflow of FUSE is presented in Algorithm C1. The full process of FUSE can be divided into three parts: initialization, three-tier U-shape split learning-based local training, and two-stage semi-asynchronous aggregation.

In the initialization phase, all parts of the model are initialized. Then the corresponding part of the model is encrypted and sent to all the TDs and CSs, respectively. Then, for each TD_i along with the corresponding CS_i, the three-tier U-shape split learning-based local training is performed. To guarantee that the feature of the consumption data and labels of consumers are maintained at the place where they are derived (data in SMs, label in TDs), a novel variant of split learning — three-tier (SM, TD, and CS) U-shape split learning is deployed in FUSE. The full theft detection model is split into the feature extractor (lower layer of the model), the feature learner (middle majority layer of the model), and the feature classifier (top layer of the model). Considering the limited resources of both SMs and TDs, assisted CSs are used to perform the majority of the model training tasks, thereby reducing the computational overhead. As shown in Figure 1, the forward propagation and the backpropagation of the training process follow the path of the green and red lines. Specifically, the forward propagation is performed as follows:

$$\mathrm{output}_{\mathrm{FE}_j}^{e-1} \leftarrow \mathrm{ForwardPropagation}(w_{\mathrm{FE}}^{e-1}, x_i^j),$$
$$\mathrm{output}_{\mathrm{FL}}^{e-1} \leftarrow \mathrm{ForwardPropagation}(w_{\mathrm{FL}}^{e-1}, \mathrm{output}_{\mathrm{FE}}^{e-1}),$$
$$\hat{y} \leftarrow \mathrm{ForwardPropagation}(w_{\mathrm{FC}}^{e-1}, \mathrm{output}_{\mathrm{FL}}^{e-1}).$$

The gradients of each component are then backpropagated. The data flow takes on the shape of the letter “U”, resulting in the adoption of the term “U-shape split learning” for this approach. Thus, by deploying the U-shape split learning, the owner of the label (TD) does not require any label sharing. In order to enhance privacy-preserving, fully homomorphic encryption is applied such that the aggregator

*Corresponding author (email: guan@ncepu.edu.cn)

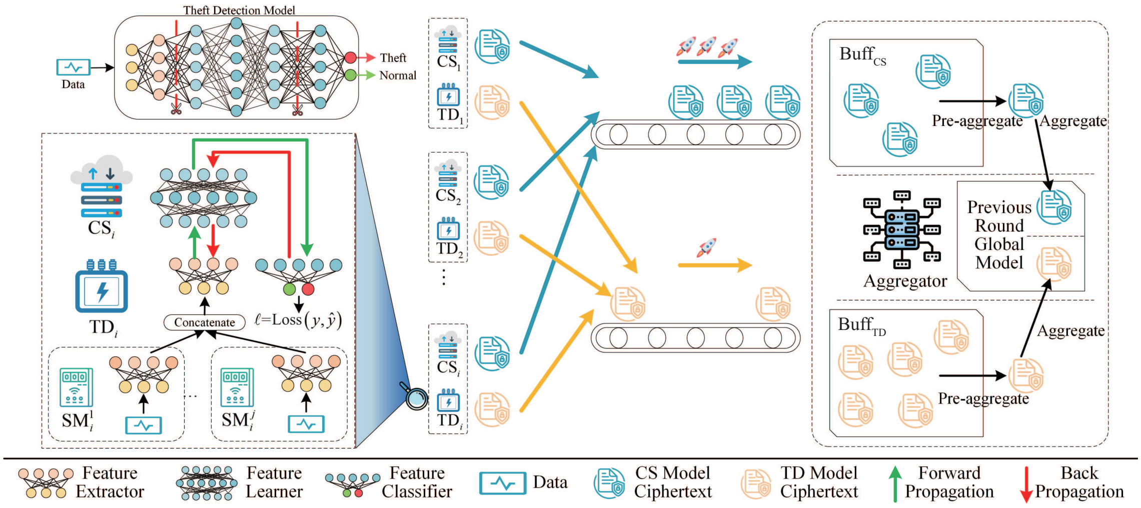

Figure 1 (Color online) Overview of the proposed framework.

is unable to infer sensitive information from the local models.

Practically, the model submission time may vary between a given TD_i and its corresponding CS_i, since CS_i updates the feature learner in advance of the update on the feature extractor by TD_i due to the order in the U-shape split learning-based local training process. Inspired by the work in [5], we address this issue by aggregating the models from TD_i and CS_i respectively, such that the CS_i is not required to wait for the TD_i during model submission.

After finishing local model training, each part of the model is sent to the aggregator respectively for two-stage semi-asynchronous aggregation. As a result of the communication distance, uplink bandwidths, and computing capacities, the submission time of the local models also varies across different TD_i and CS_i. Two buffers are set to collect the models received from TDs and CSs. The models in the same buffer may be derived from basic global models in different rounds. The aggregator first pre-aggregates the top-m models which have the highest cosine similarity denoted as cos with the global model in the previous round. This pre-aggregation is composed of deriving the weighted average parameters, the dataset size, and the staleness regarding cos, as follows: $[\![v^{\mathrm{new}}]\!]_{\mathrm{pk}} = \sum_{i,k,j\in\mathrm{Buff}} \frac{\cos^j}{\sum_{j\in\mathrm{Buff}} \cos^j}[\![v_i^k]\!]_{\mathrm{pk}}$, $|\overline{D}| = \sum_{i,j\in\mathrm{Buff}} \frac{\cos^j}{\sum_{j\in\mathrm{Buff}} \cos^j}|D_i|$, $\overline{\tau} = \sum_{k,j\in\mathrm{Buff}} \frac{\cos^j}{\sum_{j\in\mathrm{Buff}} \cos^j}(r-k)$. Then the global model is aggregated as follows:

$$[\![v^r]\!]_{\mathrm{pk}} \leftarrow (1-\theta_r)[\![v^{r-1}]\!]_{\mathrm{pk}} + \theta_r[\![v^{\mathrm{new}}]\!]_{\mathrm{pk}}, \tag{1}$$

where θ_r is the final weight of the updated model determined by the product of a hyperparameter $\theta \in [0,1]$ and $g(\cdot)$, which is a function of staleness and dataset size, i.e., $\theta_r = \theta g(\overline{\tau}, |\overline{D}|)$. $g(\cdot)$ is defined as follows:

$$g(\overline{\tau}, |\overline{D}|) = \frac{1}{\left(1+\overline{\tau}+\frac{\sum|D_i|}{M|\overline{D}|}\right)^{\frac{1}{e}}}. \tag{2}$$

The detail of FUSE is elaborated in Appendix C.

Experiment. Detailed experimental studies have been carried out to evaluate FUSE by using a real-world dataset. We first explore the hyperparameters of our framework. Then the selected hyperparameters are applied to evaluate the performance of FUSE. What is more, we analyze the communication and computation overhead of our framework. The detailed results provided in Appendix D demonstrate the improvements achieved by our framework, including a better model performance and a reduction in the computational overhead for resource-constrained devices.

Related work. The related work is provided in Appendix E.

Conclusion. In this study, we propose a novel theft detection framework named FUSE. Firstly, we introduce a new variant of split learning named three-tier U-shape split learning into the local training process. This allows us to migrate the extensive computational overhead to the assisted CSs, while ensuring the sensitive data is preserved in the place where it is generated for privacy-preserving. Furthermore, we design a two-stage semi-asynchronous aggregation mechanism to accommodate the straggler issue and associated communication overhead, which consists of cosine similarity-based pre-aggregation and staleness-aware aggregation. Finally, we conduct extensive experiments and validate our model performance through the comparisons with the benchmarks.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 62372173).

Supporting information Appendixes A–E. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Yan Z Z, Wen H. Performance analysis of electricity theft detection for the smart grid: an overview. IEEE Trans Instrum Meas, 2022, 71: 1–28
2 Su Z, Wang Y T, Luan T H, et al. Secure and efficient federated learning for smart grid with edge-cloud collaboration. IEEE Trans Ind Inf, 2021, 18: 1333–1344
3 Depuru S S S R, Wang L F, Devabhaktuni V. Electricity theft: overview, issues, prevention and a smart meter based approach to control theft. Energy Policy, 2011, 39: 1007–1015
4 McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 2017. 1273–1282
5 Chen Y, Sun X Y, Jin Y C. Communication-efficient federated deep learning with layerwise asynchronous model update and temporally weighted aggregation. IEEE Trans Neural Netw Learn Syst, 2019, 31: 4229–4238