Preliminary Results on Sensitive Data Leakage in Federated Human Activity Recognition

1st Riccardo Presotto
Dept. of Computer Science
University of Milan
Milan, Italy
riccardo.presotto@unimi.it

2nd Gabriele Civitarese
Dept. of Computer Science
University of Milan
Milan, Italy
gabriele.civitarese@unimi.it

3rt Claudio Bettini
Dept. of Computer Science
University of Milan
Milan, Italy
claudio.bettini@unimi.it

***Abstract*—Sensor-based Human Activity Recognition (HAR) has been a hot topic in pervasive computing for many years, as an enabling technology for several context-aware applications. However, the deployment of HAR in real-world scenarios is limited by some major challenges. Among those issues, privacy is particularly relevant, since activity patterns may reveal sensitive information about the users (e.g., personal habits, medical conditions). HAR solutions based on Federated Learning (FL) have been recently proposed to mitigate this problem. In FL, each user shares with a cloud server only the parameters of a locally trained model, while personal data are kept private. The cloud server is in charge of building a global model by aggregating the received parameters. Even though FL avoids the release of labelled sensor data, researchers have found that the parameters of deep learning models may still reveal sensitive information through specifically designed attacks. In this paper, we propose a first contribution in this line of research by introducing a novel framework to quantitatively evaluate the effectiveness of the Membership Inference Attack (MIA) for FL-based HAR. Our preliminary results on a public HAR dataset show how the global activity model may actually reveal sensitive information about the participating users and provide hints for future work on countering such attacks.**

***Index Terms*—human activity recognition, federated learning, privacy**

I. Introduction

Sensor-based Human Activity Recognition (HAR) enables the development of applications in several areas, including healthcare and well-being [1]. Currently, the most accurate HAR approaches rely on fully-supervised Deep Learning (DL) methods. These models are usually trained in a centralized fashion, using labeled sensor data from multiple individuals. However, activity data may include private and potentially sensitive information about the users (e.g., health conditions, habits) [2]. Federated Learning (FL) is a promising framework to mitigate privacy issues [3]. Indeed, in FL each user locally trains a personal model using the available labeled data. The personal model parameters of each participating user are forwarded to a cloud server that is in charge of aggregating them to generate a shared global model. In the last few years, several research groups have investigated FL-based solutions for HAR [4]. However, recent studies reveal that the model's parameters received and manipulated by the cloud server may still reveal sensitive information about the FL participating users [5]. Considering honest-but-curious attackers, this scenario opens up to two privacy problems: 1) a cloud service provider may infer private information for a specific user by exploiting the global model parameters or by exploiting the personal model parameters received from that participant, and 2) a participating user may infer sensitive information about another participant by attacking the global model parameters periodically received by the cloud server. While it is possible to prevent the cloud server to attack personal model parameters by using Secure Multiparty Computation [6] (i.e., by preventing the cloud server to see the individual contributions of the users), attacks on the global model remain possible.

While several research works recently investigated how to infer sensitive information from FL models in several domains, to the best of our knowledge the potential privacy leakages of federated HAR models have not been studied yet. With respect to other well-studied problems like image classification, the sensor-based HAR domain has its peculiarities. Indeed, several new open research questions arise in this area. Is it possible to understand from the global model which users participated to FL training? Is it possible to understand which activities they performed, when, where, and how?

In this work, we make a first step along this line of research by proposing a novel framework to quantitatively measure the potential information leakage of the global models' weights in federated HAR. Our framework relies on the Membership Inference Attack (MIA) [7] to infer the following sensitive information about participating users: a) whether a specific subject is one of the FL participating users, b) whether a specific participant contributed to the global model with a particular activity. Our preliminary experimental evaluation suggests that it is possibile to derive sensitive information from HAR global models. Hence, we hope that this work may pave the way to further research investigations in this area.

The contributions of this work can be summarised as follows:

- We investigate the potential privacy problems in FL-based HAR, proposing research directions in this area.
- We introduce a novel framework based on the Membership Inference Attack to quantitatively measure privacy leakages of the global model in FL-based HAR.

- Our preliminary results on a public dataset suggest that the global activity model may reveal sensitive high-level information from participating users and provide hints for future work on countering such attacks.

II. Related work

Federated Learning (FL) has been proposed as a privacy-preserving framework for distributed machine learning to mitigate the privacy issues of centralized solutions [3], [8]. In FL, each client trains a local model with its available labeled data, and it only transmits the model parameters to the server instead of data. The server aggregates the received parameters and generates the global DL model accordingly. Recently, several FL solutions have been proposed for the HAR domain [4], [9], [10].

However, recent studies show that even the parameters of DL models may reveal sensitive information about training data. In the literature, the most investigated attack techniques are: a) the Membership Inference Attack (MIA) [7], [11], [12], b) the Property Inference Attack (PIA) [13], and c) the Reconstruction Attack (RA) [14]–[16]. Since our framework is based on MIA, we describe this technique in detail in Section III. The PIA technique aims at extracting properties of training data that may not be directly related to the task of the classifier (e.g., using the HAR model to infer if a subject suffers from the Parkinson disease). The RA technique aims at reconstructing prototypical examples of the samples used to train the machine learning model (e.g., reconstructing sensor data patterns that reveal sensitive physical characteristics of a subject).

These attacks may be countered by adopting additional privacy-preserving mechanisms like Differential Privacy (DP) [17], [18] and Secure MultiParty Computation (SMC) [19]. However, those approaches provide privacy at the cost of a reduced classification rate or system efficiency degradation. Understanding and balancing those trade-offs is one of the major challenges in this area.

III. Membership Inference Attack in FL-based HAR

A. Membership Inference Attack

The objective of the Membership Inference Attack (MIA) is to infer whether a specific data sample has been used or not to train a DL model. Formally, given a set X of data samples (represented by feature vectors), let D^t be a labeled dataset of pairs (x, y) where $x \in X$ and y is a label. D^t is used to train a target model F^t.

MIA assumes the access to F^t and uses a binary classifier (i.e., the attack model) to determine if a data sample $x \in X$ appears in a pair (x, y) of D^t or not. In the first case we say that x is a *member* data sample, while in the second case is a *non-member*. The attack model performs such classification by analyzing the behaviour [1] of F^t in classifying the feature vector x. Details about the construction of the attack model will be given in Section IV-A considering the specific domain of HAR.

B. Membership Inference Attack in Federated Learning

In FL, an attacker may perform the MIA attack on the global model (the target model F^t). Since the FL cloud service provider has no access to the training dataset D^t, the authors in [11], [12] proposed to train the attack model using a *shadow model* trained with a *shadow training dataset*.

A *shadow model* F^s aims at imitating the behaviour of F^t. In particular, the attacker creates a pair of disjoint *shadow training sets* D^s (members shadow data) and N^s (non-members shadow data), such that each training set contains labelled data samples in the same feature and label space as D^t. Moreover, these training datasets should have a similar distribution to D^t. In practice, shadow training datasets can be obtained by public datasets or by generating synthetic data.

F^s is trained by using D^s, and the attack model is trained by analysing the behaviour of F^s while classifying the data samples in D^s and N^s. The intuition is that, since both F^t and F^s are trained with data that share a similar data distribution, the attack model trained considering the behaviour of F^s in classifying members and non-members data samples would also be effective for F^t.

C. Shadow models for HAR

Considering the specific HAR domain, the generation of a shadow dataset D^s is particularly challenging. This is a well-known limitation of the attacks based on MIA: approximating the distribution of data strictly related to a specific set of individuals is challenging [20]. In HAR, due to the high intra- and inter-variability in activity execution among several subjects (i.e., each subject has peculiar activity patterns and habits), the underlying data distribution is not independent and identically distributed (non-IID). If D^s is significantly different from D^t, the attack performance of MIA degrades accordingly [21]. This problem becomes serious when D^t includes a large number of users with different characteristics.

For the sake of this work, we use MIA as a tool to quantify the private information that can be potentially inferred from the global model. For this reason we consider a worst-case scenario in which the attacker manages to use a shadow dataset D^s very close[2] to the actual training dataset D^t. Moreover, similarly to other applications of the MIA, we assume that the attacker has access to some data samples of the participating users to perform the attack.

IV. Our framework

In the following, we propose a novel framework based on Membership Inference Attack (MIA) to quantitatively measure the amount of sensitive information potentially revealed by the global model in FL-based HAR. Our framework relies on MIA to derive high-level properties about specific subjects from the global model. In particular, we investigate two research questions:

[1] Examples of relevant behaviours are the gradients variations and the confidence of the model while classifying an input data.

[2] In the experiments this is implemented by taking $D^s \subset D^t$.

1) *User Membership*: Is it possible to infer from the global model whether a certain user took part in the FL process? This property may be crucial considering FL systems that are specialised for a certain category of users (e.g., subjects with the same disease).
2) *Activity Membership*: Is it possible to infer from the global model whether a participating user performed a specific activity?

For the sake of this work, we only consider honest-but-curious attackers that infer sensitive data by periodically observing the parameters of the global model: the *cloud server* and the *participating users*.

A. Attack model training

Considering the notation introduced in Section III, in our setting $D^t = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ is the set of labeled samples from all the participating clients, while F^t is the global model on the cloud server. F^t is trained with a FL approach. In order to perform the MIA attack, the attacker trains a binary classifier A to determine if a given data sample belongs to D^t. In particular, we take advantage of the attack model recently proposed in [7]. This attack assumes that the attacker can inspect the internal parameters of F^t. In our FL setting, this is actually possible. Figure 1 depicts a high-level data-flow of the attack model training. In order to train A, the attacker creates the shadow datasets D^s and N^s, as well as a shadow model F^s trained using D^s. We recall that D^s has a similar distribution to D^t. Then, each data sample in D^s and in N^s is provided to F^s for classification. While processing each input, the attacker observes the behavior of F^s. In particular, given an input x provided to the shadow model F^s, the attacker extracts:

- The confidence of F^s in classifying x
- The output of each layer of F^s while processing x
- The classification loss $\ell(F^s(x), y)$
- The gradients of the loss with respect to each parameter of F^t

These values are encoded in a feature vector, that is labeled as *member* if $x \in D^s$ and *non-member* if $x \in N^s$. The resulting labeled feature vectors are used to train A.

B. Inferring user and activity membership

In the following, we illustrate how our framework infers user and activity membership using an attack model based on MIA.

Let $U = \{u_1, \ldots, u_n\}$ be a set of n users. In order to answer the research question 1), we use the attack model to infer whether a certain user $u \in U$ contributed in training the global model. In this scenario, we assume that the attacker knows the corresponding user for each available data sample. The attacker infers that a subject u participated in training the global model if the majority of the samples of u tested by the attacker are classified as *members*. We quantitatively estimate the success of the attack by computing the average confidence of the attack model in classifying data samples of u as *members*.

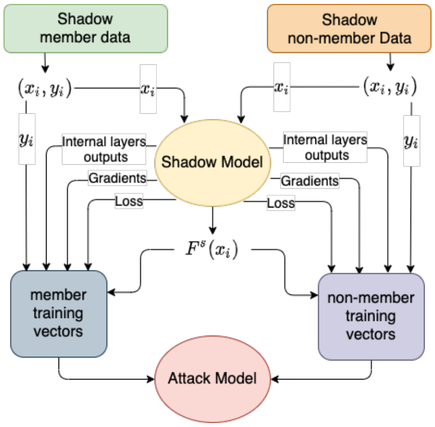


Fig. 1: Training of the attack model. The attacker observes the behavior of the shadow model when classifying *member* and *non-member* data points. The output is the training dataset for the *attack model*.

In order to answer the research question 2), we use the attack model to infer whether u participated in training the global model with an activity a. In this scenario, we also assume that the attacker knows the activity label for each available sample. The attacker infers that u participated in training the global model with activity a when the majority of the available labeled samples of u related to the activity a tested by the attacker are classified as *members*. We quantitatively estimate the success of the attack by computing the average confidence of these classifications.

V. Experimental evaluation

A. Dataset

We perform a preliminary evaluation of our framework using the publicly available MobiAct dataset [22]. In particular, MobiAct includes labeled data from inertial sensors (i.e., accelerometer, gyroscope, and magnetometer) from a smartphone placed in the pant's pocket. Overall, MobiAct includes data from 60 subjects. In our experiments, we considered the following physical activities [3]: *standing*, *walking*, *jogging*, *jumping*, and *sitting*. Since this dataset involves a relatively large number of subjects with respect to other sensor-based HAR datasets, it is particularly suited to evaluate FL-based solutions. In our experiments, we consider a FL client for each user in MobiAct.

B. Experimental setup

1) Federated Learning: We use the FL experimental setup recently proposed in [23], since it exhibited promising performances for HAR. In particular, the activity model is a feed-forward deep neural network composed of three fully connected layers having respectively 128, 64, and 32 neurons,

[3] Note that we omitted from MobiAct those physical activities with a limited number of samples as they are insufficiently represented and hence not suitable for our evaluation. We believe that this problem is only related to this specific dataset and that, in realistic settings, even short activities would be represented by a sufficient number of samples.

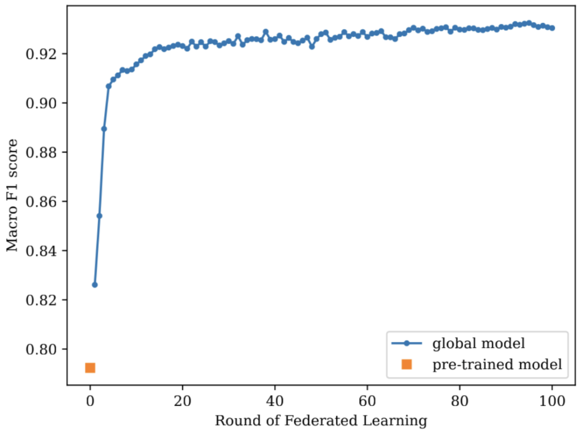


Fig. 2: The evolution of the FL-score at different communication rounds

and a softmax layer for classification. The inputs of that network are hand-crafted feature vectors extracted in real-time from the stream of sensor data. We consider features that proved to be effective for HAR in the literature [24]. We used Adam [25] as optimiser. The well-known FedAvg algorithm [3] is in charge of aggregating the model parameters received by clients and updating the global model. Each client trains its local model for 10 epoch. Finally, we empirically selected 30 as the number of FL communication rounds as it guarantees the convergence of the *global model* avoiding overfitting (see Figure 2).

2) Membership Inference Attack: The implementation of MIA is based on the public *ML Privacy Meter*[4] tool [7]. For each experiment, we trained the *attack model* for 150 epochs with a learning rate of 0.001, while the Adam optimiser was used to minimise the loss function. As we mentioned in Section III-C, in our experiments the shadow model is trained by using a subset of labeled data from the participating users.

3) Metric: In order to quantitatively measure the probability that a sample x was part of the target dataset D^t given a target model F^t, we use the confidence of the attack model in classifying x as *member*. We will refer to this measure as the *membership probability* (MP):

$$MP(x) = Pr(x \in D^t | F^t)$$

Intuitively, an MP value closer to 1 indicates that x is likely a *member*, while an MP value closer to 0 indicates that x is likely a *non-member*.

4) Recogniton rate: Before evaluating our framework, we performed an initial experiment to evaluate the recognition rate of FL on MobiAct. The evaluation was performed considering 60 clients (one for each user). For each client, we used 70% of data to train the FL model and the remaining 30% for testing. Figure 2 shows the outcome of our experiment. From these results we observed that the classifier quickly converges to high F-1 scores.

C. Evaluating user membership

1) Data preparation: The data partitioning schema is illustrated in Figure 3. As usually proposed in FL methods, we randomly select 15% of the users from the dataset to initialize the global model (pre-training). The remaining users are partitioned as follows: 50% of users participate to FL (FL members) and 50% of users do not participate to FL (FL non-members) [5]. The global model is hence trained in a FL fashion using data in D^t. We train the attack model by using 70% of data from D^t labeled as *members*, and 70% from labeled as *non-members* [6]. We use the remaining 30% from both datasets to evaluate the effectiveness of the attack model.

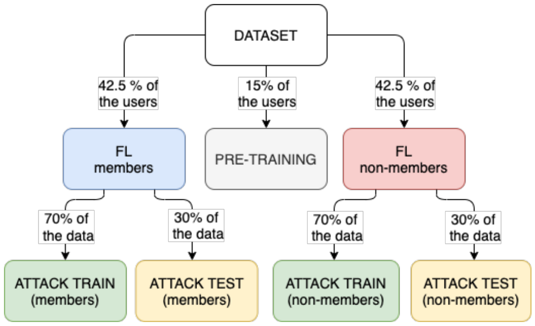


Fig. 3: Dataset splitting process adopted to evaluate user membership

2) Results: Figure 4 shows the results of the user membership attack at the data sample granularity. We observed a MP value close or equal to 1 for most of the *FL members*' data samples, while a value close or equal to 0 for most of the *FL non-members*' samples. Thus, we can conclude that, overall, the *attack model* is confident in discriminating *members* and *non-members* samples.

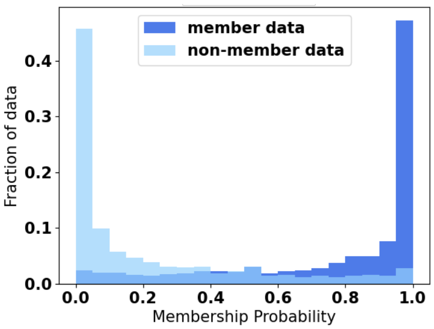


Fig. 4: Distribution of the membership probability for *members* versus *non-members* data

Figure 5 shows the same result at the user granularity. In particular, for each user we average the MP score computed on its test data sample. We can observe that the users that actually participated in FL are associated with an average higher MP value than those that did not participate. Hence, in

[4] https://github.com/privacytrustlab/ml_privacy_meter

[5] Note that the union of labeled data from *FL members* corresponds to D^t.

[6] Note that these partitions correspond to D^s and N^s, respectively.

this scenario, the MIA attack potentially reveals if a specific user participated to FL.

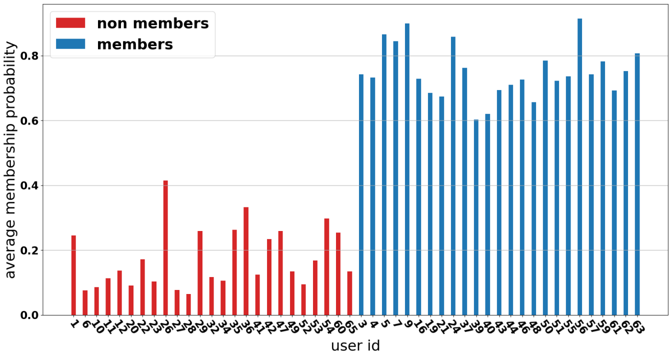


Fig. 5: Average membership probability assigned by the attack model to each of the considered users.

D. Evaluating user membership with data not used in FL

In this experiment, we want to check if the attack recognises the membership of a user even by analysing data samples from that user that have not been used in training the global model.

1) Data preparation: In order to perform this experiment, we consider the specific setting where the attacker has access to 15% of data samples (not used to train the global model) from 5% of the FL members. The data partitioning schema is depicted in Figure 6.

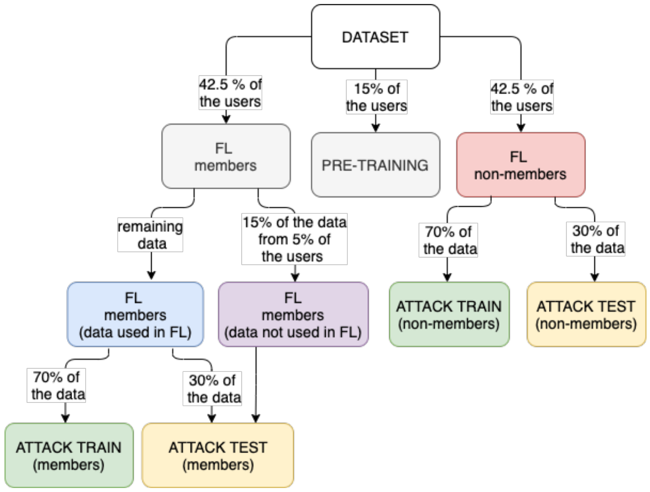


Fig. 6: Dataset splitting process adopted to evaluate user membership with data not used in FL

2) Results: Figure 7 summarizes the results of the attack at the user granularity. We observed that data samples not used in the FL training still reliably reveal the membership of the corresponding users.

E. Evaluating activity membership

In this experiment, we consider the setting proposed in Section V-C to understand if it is possible to determine whether a user contributed to FL with a specific activity.

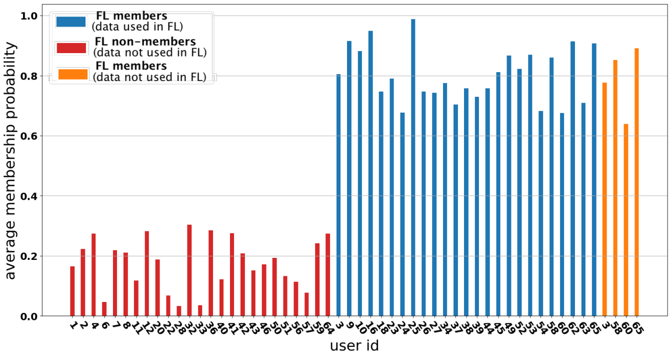
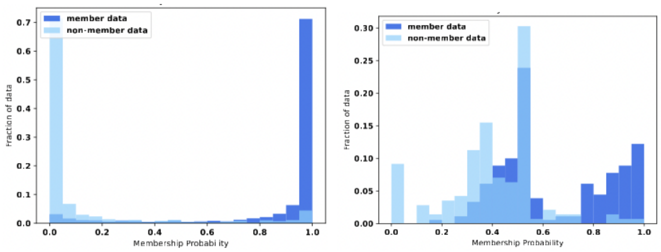


Fig. 7: Average membership probability assigned by the attack model to each of the considered users.

For each activity, we computed the MP value for each test data sample of both FL members and non-members subjects. Figure 8 shows the outcome of this experiment considering the activities *walking* and *sitting*.



(a) MP Walking

(b) MP Sitting

Fig. 8: MP assigned to the samples of the activities *Walking* and *Sitting*

We observe that the attack is effective for *walking* (a) while not really for *sitting* (b). Indeed, Figure 8b shows that the average MP is around 0.5. Since both FL members and non-members perform this activity in a similar way, during the training phase the attack model can not observe significant differences in the shadow model behaviour when processing members and non-members data.

Intuitively, considering the sensor setup used in MobiAct and in several other HAR datasets, *walking* represents a set of activities that are likely to differ in their pattern of execution by different subjects, while *sitting* represents activities that have limited variance in their execution patterns.

This may lead to conclude that the attack for this last category of activities is not effective while it is effective for those in the first category.

Nonetheless, activities in this first category are not necessarily exposed to privacy risks in general. Indeed, considering larger datasets where it is very unlikely that users perform the activity in a unique way, it is questionable if the attack would be effective as well.

Considering possible privacy protection approaches, we believe that these results may provide useful information on which activities may be more exposed, hence guiding, for example, the distribution of artificial noise in obfuscation strategies.

VI. Conclusion and future work

In this work, we investigated the problem of measuring the potential privacy leakages of FL-based HAR models. In particular, we proposed a novel framework based on the Membership Inference Attack. Our preliminary results suggest that the global activity model may actually reveal sensitive information about the participating users.

However, this is only the very first step of a research direction that we intend to explore in the near future. For instance, we want to investigate whether the global model parameters can also reveal sensitive information related to HAR, including when, where, and how a user performed a specific activity.

A major limitation of this work is using a subset of the target data to train the shadow model. Clearly, this is not realistic since the attacker cannot actually access this information. We will investigate alternative strategies to train the shadow model (e.g., using GAN to generate synthetic data) as well as unsupervised membership attack methods [7], [26].

We also plan to evaluate other types of attacks besides MIA. For instance, the reconstruction attack may be used to recreate sensor patterns that reveal medical conditions of the participating users, while the property-inference attack could be used to infer high-level properties about specific users from the global activity model.

Considering privacy preserving techniques, we plan to study solutions based on Local Differential Privacy (LDP) with heuristics guided by MIA-based analysis as mentioned in the experimental section.

Finally, we will also consider additional HAR datasets to more robustly assess our framework.

References

[1] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012.

[2] C. Bettini and D. Riboni, "Privacy protection in pervasive systems: State of the art and technical challenges," *Pervasive and Mobile Computing*, vol. 17, pp. 159–174, 2015.

[3] 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.

[4] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "Fedhealth: A federated transfer learning framework for wearable healthcare," *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 83–93, 2020.

[5] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, and H. Qi, "Beyond inferring class representatives: User-level privacy leakage from federated learning," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 2512–2520.

[6] Y. Li, Y. Zhou, A. Jolfaei, D. Yu, G. Xu, and X. Zheng, "Privacy-preserving federated learning framework based on chained secure multiparty computing," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6178–6186, 2020.

[7] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning," in *2019 IEEE Symposium on Security and Privacy (SP)*, 2019, pp. 739–753.

[8] L. Lyu, X. He, Y. W. Law, and M. Palaniswami, "Privacy-preserving collaborative deep learning with application to human activity recognition," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1219–1228.

[9] K. Sozinov, V. Vlassov, and S. Girdzijauskas, "Human activity recognition using federated learning," in *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)*. IEEE, 2018, pp. 1103–1111.

[10] X. Ouyang, Z. Xie, J. Zhou, J. Huang, and G. Xing, "Clusterfl: a similarity-aware federated learning system for human activity recognition," in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 2021, pp. 54–66.

[11] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *2017 IEEE Symposium on Security and Privacy (SP)*, 2017, pp. 3–18.

[12] S. Truex, L. Liu, M. E. Gursoy, L. Yu, and W. Wei, "Demystifying membership inference attacks in machine learning as a service," *IEEE Transactions on Services Computing*, pp. 1–1, 2019.

[13] 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)*, 2019, pp. 691–706.

[14] B. Hitaj, G. Ateniese, and F. Pérez-Cruz, "Deep models under the gan: Information leakage from collaborative deep learning," *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017.

[15] Y. Zhang, R. Jia, H. Pei, W. Wang, B. Li, and D. Song, "The secret revealer: Generative model-inversion attacks against deep neural networks," 06 2020, pp. 250–258.

[16] L. Zhu, Z. Liu, and S. Han, *Deep Leakage from Gradients*. Red Hook, NY, USA: Curran Associates Inc., 2019.

[17] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," 2018.

[18] G. Cormode, S. Jha, T. Kulkarni, N. Li, D. Srivastava, and T. Wang, "Privacy at scale: Local differential privacy in practice," in *Proceedings of the 2018 International Conference on Management of Data*, ser. SIGMOD '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1655–1658. [Online]. Available: https://doi.org/10.1145/3183713.3197390

[19] 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*. New York, NY, USA: Association for Computing Machinery, 2017, p. 1175–1191. [Online]. Available: https://doi.org/10.1145/3133956.3133982

[20] 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.

[21] A. Salem, Y. Zhang, M. Humbert, P. Berrang, M. Fritz, and M. Backes, "Ml-leaks: Model and data independent membership inference attacks and defenses on machine learning models," *arXiv preprint arXiv:1806.01246*, 2018.

[22] C. Chatzaki, M. Pediaditis, G. Vavoulas, and M. Tsiknakis, "Human daily activity and fall recognition using a smartphone's acceleration sensor," 07 2017, pp. 100–118.

[23] C. Bettini, G. Civitarese, and R. Presotto, "Personalized semi-supervised federated learning for human activity recognition," *arXiv preprint arXiv:2104.08094*, 2021.

[24] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recogn. Lett.*, vol. 119, pp. 3–11, 2019.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[26] Y. Bai, D. Chen, T. Chen, and M. Fan, "Ganmia: Gan-based black-box membership inference attack," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.