Privacy-Preserving Federated Learning

Federated learning is a decentralized machine learning approach where the data remains in the parties' premises and the model is trained via intermediate gradients. By default, federated learning does not provide privacy due to several privacy attacks such as membership inference or data reconstruction attacks [1-4]. The aim of this project is to choose two privacy-preserving federated learning frameworks that employ homomorphic encryption, secure multiparty computation, or differential privacy and compare their results in terms of efficiency, utility, and privacy. At the end of this project, the students are expected to come-up with a novel approach for privacy-preserving federated learning.

Privacy Attacks on Federated Learning

Federated learning is a decentralized machine learning approach where the data remains in the parties' premises and the model is trained via intermediate gradients. By default, federated learning does not provide privacy due to several privacy attacks such as membership inference or data reconstruction attacks [1-4]. The aim of this project is to re-implement and improve state-of-the-art attacks on federated learning by using real-life datasets such as biomedical data or smart-meter data. Second track for this project could be the to re-implement and improve state-of-the-art attacks on non-iid data distributions.

[1] B. Hitaj, G. Ateniese, and F. Perez-Cruz. Deep models under the GAN: Information leakage from collaborative deep learning. In ACM CCS, 2017.

[2] 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.

[3] L. Zhu, Z. Liu, and S. Han. Deep leakage from gradients. In NeurIPS, volume 32. 2019.

[4] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov. Exploiting unintended feature leakage in collaborative learning. In IEEE S&P, 2019.

Mitigating machine learning attacks

There are various attacks to machine learning models such as model stealing, data reconstruction, and membership inference attacks. Follow the state-of-the-art machine learning attacks, choose one specific attack and propose a novel defense mechanism. You should compare your defense mechanism with the proposed defenses (at least two) in the literature.

Some example papers:

[1] Salem, Ahmed, et al. "Ml-leaks: Model and data independent membership inference attacks and defenses on machine learning models." *arXiv preprint arXiv:1806.01246* (2018).

[2] Li, Zheng, and Yang Zhang. "Membership leakage in label-only exposures." *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. 2021.

[3] Jia, Jinyuan, et al. "Memguard: Defending against black-box membership inference attacks via adversarial examples." *Proceedings of the 2019 ACM SIGSAC conference on computer and communications security*. 2019.

Phenotype – Genotype Inference

Use machine-learning algorithms to quantify the risk of phenotype inference from genomes of individuals. Use opensnp dataset (which has all genotype and phenotype information). Quantify the risk for different scenarios.

Data collection, analysis, and risk assessment for profile matching on social networks

Identify the risk of profile matching in various social networks. Come up with data collection strategies. Explore the risk using collected data.

Nowadays online social networks (OSNs) have become so popular. The tremendous amount of data that these online social networks contain can reveal relevant information about their users. However, whenever someone wants to analyze OSNs, the first problem they encounter is how to find real data with ground truths. By ground truth, we mean that a user has accounts in both OSNs and we know both of these accounts. In this project, you will explore different scraping (crawling) techniques and crawl data from two OSNs of your choice. Note, the goal here is to collect as many ground truths as possible.

Clustering in Online Social Networks

In today's world, online social networks (OSNs) have become a popular way of communication. The analysis of this enormous amount of data that is accumulating in these networks plays an important role in different scientific areas. Clustering the networks does not only reduce the complexity of analyzing large OSNs, it can also help us understand the dynamics of the clustered communities. In this project, you will classify two online social networks by using different clustering algorithms and analyze the results. Part of this analysis will be the observation of two nodes of a community in one network and their relationship in the other network. You need to have some ground truths in order to perform the analysis.

Clustering-based Profile Matching

In today's world, online social networks (OSNs) have become a popular way of communication. The tremendous amount of data shared in these social networks can reveal relevant information about their users. Therefore, it is of significant interest to measure the privacy leak coming due to profile matching. There exist many works on profile matching based on publicly available attributes such as user name, location, profile photo, description, number of connections etc. They mostly suffer from high complexity considering that the number of users in OSNs is increasing every day. In this project, you will match user profiles across online social networks, by first clustering the users in each social network in order to reduce the number of possible matches for each target user and quantify its privacy leakage.

Some previous works on profile matching:

A. A. Malhotra, L. C. Totti, W. M. Jr., P. Kumaraguru, and V. Almeida. Studying user footprints in different online social networks. In ASONAM, 2012.

B. R. Zafarani and H. Liu. Connecting users across social media sites: A behavioral-modeling approach. In KDD, 2013.
C. S. Nilizadeh, A. Kapadia, and Y.-Y. Ahn. Community-enhanced de-anonymization of online social networks. In CCS, 2014.

Mitigation Techniques for Profile Matching Attacks

Matching user profiles across online social networks is useful for online service providers to build complete profiles of individuals (e.g., to provide better personalized advertisement), but at the same time, it also has serious privacy concerns. In this project, you will provide recommendations to the users (about the content they share) to reduce their risk to profile matching attacks. Such recommendations may include (i) generalizing or distorting some shared content of the user (e.g., generalizing the shared location or posting a content at a later time); or (ii) choosing not to share some content (especially for attributes that are hard to generalize or distort, such as interest or sentiment inferred from users' posts).

Inferring Attributes on Online Social Networks

Previous works have proposed privacy attacks to infer attributes (e.g., locations, occupations, interests, friends etc.) of online social network users based on publicly available information provided by the users in online social networks (OSNs). They have shown that private attributes can be inferred from users' publicly available data in an accurate way. Generally, the information provided in one OSN is used to infer an attribute of a user. In this project, you will study the impact of profile matching (matching accounts belonging to the same individual in two different OSNs) to attribute(s) inference.

Some previous works on attribute inference:

A. N.Z. Gong, and B. Liu. You are who you know and how you behave: Attribute inference attacks via users' social friends and behaviors. In USENIX 2016.
B. M. Backes, M. Humbert, J. Pang, Y. Zhang. walk2friends: Inferring social links from mobility profiles. In CCS 2017.

Quantifying Privacy Leakage from Workflow Provenance and Metadata

In the modern era, especially in scientific research, it is becoming crucial to provide provenance for reproducibility and auditability purposes. Workflow systems model and record provenance as a directed acyclic graph that describes the steps performed to obtain the intermediate and final results of a computation. While sharing provenance brings numerous benefits, it may also disclose sensitive information (e.g., about the data involved in the workflow or the workflow itself). In this project, you will quantify the information leakage of provenance graphs (workflows) with their associated metadata. You will mostly focus on inferring private (hidden) attribute(s) of a target record given the provenance

graph, and its associated metadata/outputs on two use-cases: (1) predicting the presence of a disease based on physical conditions, and (2) performing genome-wide association (GWAS) meta-analysis, which are examples of studying provenance in scientific research.

Location Privacy

Location-based service providers collect location information from users to provide location-based services. To protect the privacy of users from these service providers several approaches have been proposed in the literature. One of them is adding controlled noise to data by satisfying geo-indistinguishability [1]. Another approach is satisfying k-anonymity in data sharing [2]. The aim of this project is implementing and comparing these two approaches in terms of privacy and utility.

[1] Andrés, Miguel E., Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. "Geo-indistinguishability: Differential privacy for location-based systems." In Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security, pp. 901-914. 2013.

[2] Gedik, Bugra, and Ling Liu. "Location privacy in mobile systems: A personalized anonymization model." In 25th IEEE International Conference on Distributed Computing Systems (ICDCS'05), pp. 620-629. IEEE, 2005.