Offer #2023-06241

PhD Position F/M Alternative approaches for privacy-preserving in federated learning

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

Inria is a national research institute dedicated to digital sciences that promotes scientific excellence and transfer. Inria employs 2,400 collaborators organised in research project teams, usually in collaboration with its academic partners.
This agility allows its scientists, from the best universities in the world, to meet the challenges of computer science and mathematics, either through multidisciplinarity or with industrial partners.
A precursor to the creation of Deep Tech companies, Inria has also supported the creation of more than 150 start-ups from its research teams. Inria effectively faces the challenges of the digital transformation of science, society and the economy.

Assignment

Assignments :

Context:
Federated learning (FL) enables a large number of IoT devices (mobiles, sensors) to cooperatively to learn a global machine learning model while keeping the devices' data locally [MMR+17, LSTS20]. For example, Google has applied FL in their application Gboard to predict the next word the users would type on their smartphones [HRM+18]. FL can help to mitigate privacy concerns, as the raw data is kept locally by the users and never needs to be sent elsewhere. However, maintaining the data locally does not provide itself formal privacy guarantees.

Many attacks have shown the vulnerability of federated learning systems: the adversary can reconstruct private data points (e.g., images and private features) [ZLH19, GBDM20, DXN+22], infer the membership of the data instance [MSDCS19, ZXN21] and reconstruct the local model of the user [XN21] just by eavesdropping the exchanged messages. As a result, differentially private (DP) algorithms [MRTZ18, BGTT18] have been proposed for FL to protect privacy by injecting random noise into the transmitted messages. DP ensures that if the user changes one training sample, the adversary does not observe much difference in the exchanged messages and then may not confidently draw any conclusions about the presence or absence of a specific data sample. Therefore, attacks are less efficient [JE19]. However, the noise typically deteriorates the performance of the model.

Alternatively, some methods that were initially designed to improve model generalization have been empirically shown to be effective against privacy attacks as well, as the resulting model memorizes less the training samples. For example, in the centralized training scenario, pruning the neural network [HPTD15] can mitigate the privacy leakage from membership inference [WWW+21] and model inversion [HSR+20] attacks. Mixing up training data samples [ZCDL18] may also help to defend against adversarial attacks [PXZ20]. Besides, methods which exploit other sources of randomness, like batch sampling [HT19] and mixing up the average weights in decentralized learning [XD21], can amplify the DP guarantees. However, how to adapt and combine these techniques in federated system where the devices may exhibit different computation/memory capacities and data distributions, as well as have different privacy requirements, is still an open problem.

Research Goal:
The goal of this PhD is to propose new privacy-preserving methods for FL which do not necessarily add synthetic noise to updates (as DP does), but either rely on different approaches, like parameter pruning or quantization, or exploit other sources of randomness, like the batch sampling procedure or the "mixup" procedure. Pruning and quantization can lead to models with better generalization that are in turn less vulnerable to attacks, but also reduce the computation/communication requirements of FL training and then potentially its energy consumption. Because of devices' heterogeneity (in terms of computation, memory, data, and privacy requirements), we need to design device-aware strategies, e.g., to select the level of model compression through pruning or quantization. The candidate will also study how these alternative techniques may be combined with more traditional DP approaches, potentially leading to improved accuracy-utility trade-offs against FL privacy attacks.

Reference:

[BGTT18] Aur □elien Bellet, Rachid Guerraoui, Mahsa Taziki, and Marc Tommasi. Personalized and private peer-to-peer machine learning. In Amos J. Storkey and Fernando P □erezCruz, editors, International Conference on Artificial Intelligence and Statistics, AISTATS 2018.
[DXN+22] Ilias Driouich, Chuan Xu, Giovanni Neglia, Frederic Giroire, and Eoin Thomas. A novel model-based attribute inference attack in federated learning. In FL-NeurIPS'22-Federated Learning: Recent Advances and New Challenges workshop in Conjunction with NeurIPS
[GBDM20] Jonas Geiping, Hartmut Bauermeister, Hannah Dr □oge, and Michael Moeller. Inverting gradients–how easy is it to break privacy in federated learning? NIPS, 2020.
[HPTD15] Song Han, Jeff Pool, John Tran, and William Dally. Learning both weights and connections for efficient neural network. Advances in neural information processing systems, 28, 2015.
[HRM+18] Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Fran □coise Beaufays, Sean Augenstein, Hubert Eichner, Chlo □e Kiddon, and Daniel Ramage. Federated learning for mobile keyboard prediction. arXiv preprint arXiv:1811.03604, 2018.
[HSR+20] Yangsibo Huang, Yushan Su, Sachin Ravi, Zhao Song, Sanjeev Arora, and Kai Li. Privacy-preserving learning via deep net pruning. arXiv preprint arXiv:2003.01876, 2020.
[HT19] Stephanie L Hyland and Shruti Tople. An empirical study on the intrinsic privacy of stochastic gradient descent, 2019. arXiv:1912.02919.
[JE19] Bargav Jayaraman and David Evans. Evaluating differentially private machine learning in practice. In USENIX Security Symposium, 2019.
[LSTS20] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. IEEE signal processing magazine, 37(3):50–60, 2020.
[MMR+17] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agueray Arcas. Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics, pages 1273–1282. PMLR, 2017

[MRTZ18] H. Brendan McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang. Learning differentially private recurrent language models. In 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018,
[MSDCS19] Luca Melis, Congzheng Song, Emiliano De Cristofaro, and Vitaly Shmatikov. Exploiting unintended feature leakage in collaborative learning. In 2019 IEEE Symposium on Security and Privacy (SP), pages 691–706. IEEE, 2019.
[PXZ20] Tianyu Pang, Kun Xu, and Jun Zhu. Mixup inference: Better exploiting mixup to defend adversarial attacks. In 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net, 2020.
[WWW+21] Yijue Wang, Chenghong Wang, Zigeng Wang, Shanglin Zhou, Hang Liu, Jinbo Bi, Caiwen Ding, and Sanguthevar Rajasekaran. Against membership inference attack: Pruning is all you need.,IJCAI 2021
[XD21] Hanshen Xiao and Srinivas Devadas. Towards understanding practical randomness beyond noise: Differential privacy and mixup. Cryptology ePrint Archive, 2021.
[XN21] Chuan Xu and Giovanni Neglia. What else is leaked when eavesdropping federated learning? In CCS workshop Privacy Preserving Machine Learning (PPML), 2021.
[ZCDL18] Hongyi Zhang, Moustapha Ciss □e, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In 6th International Conference on Learning Representations, ICLR 2018
[ZLH19] Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. In Advances in Neural Information Processing Systems, pages 14774–14784, 2019.
[ZXN21] Oualid Zari, Chuan Xu, and Giovanni Neglia. Efficient passive membership inference attack in federated learning. NeurIPS PriML workshop, 2021

Main activities

Research

Skills

The candidate should have good programming skills and previous experience with PyTorch or Tensor-Flow. He/She should also be knowledgeable on machine learning and have good analytical skills. We expect the candidate to be fluent in English.

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Contribution to mutual insurance (subject to conditions)

Remuneration

Gross Salary per month: 2051€ brut per month (year 1 & 2) and 2158€ brut per month (year 3)

General Information

- **Theme/Domain :** Security and Confidentiality
  System & Networks (BAP E)
- **Town/city :** Sophia Antipolis
- **Inria Center :** Centre Inria d'Université Côte d'Azur
- **Starting date :** 2023-09-01
- **Duration of contract :** 3 years
- **Deadline to apply :** 2023-09-30

Contacts

- **Inria Team :** COATI
- **PhD Supervisor :**
  Xu Chuan / chuan.xu@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.