# Data Privacy in Smart Cities – Federated Learning to the Rescue?

by Anastasia Pustozerova and Rudolf Mayer (SBA Research)

***Within any smart system, data is vital for making the management of resources and assets more efficient. At the same time, data is a potential vulnerability to data owners, and it could become a threat in the hands of an adversary. Data security and privacy are therefore critical for building sustainable smart systems like smart cities. In such systems, where data collection is distributed, federated learning seems like a prime candidate to address the issue of data privacy. However, there are still concerns that need to be addressed regarding privacy and security in federated learning.***

Machine learning demands large amounts of data to build effective models that can help to improve services. In many real-world scenarios, data originates at the edge, e.g., smart meters and sensors in smart power grids. In traditional machine learning workflows, data must be centralised from different sources before performing the model training. Concentrating all data in one place creates a single point of failure – an adversary that can potentially gain access to this centralised data is a threat to multiple entities.

Federated learning enhances data privacy in machine learning by suggesting a new perspective on applying machine learning for the analysis of distributed data. The main idea is to train machine learning models closer to the place where data originates – and just aggregate these trained models instead of the (sensitive or private) data. Federated learning, therefore, eliminates the need to share and centralise sensitive data, allowing data owners to keep it private while at the same time offering comparable effectiveness of models.

Federated learning architectures often consist of data owners (clients), which perform local training of the models on their own data, and a central aggregator, which collects the models from the clients and averages them, producing a global model. The global model can be sent back to the clients for the next cycle of training to improve its effectiveness, and later utilised for predictions. Some of the main challenges of federated learning include communication costs, data and systems heterogeneity. Many works propose different optimisation algorithms to tackle these issues, e.g., via client sampling or model and gradient compression [1]. However, comparatively little attention has been put on remaining privacy and security risks, and new attack vectors open up simply due to the distributed nature of federated learning (see Figure 1).

Security risks (integrity and availability). Malicious participants of federated learning or adversaries leveraging transferred information can corrupt the learning process to degrade the global model quality or to make it perform target misclassification. In smart cities, successfully executed attacks can result in adversaries manipulating situations to favour them – for example, by manipulating demand-driven pricing – or can even result in the failure of critical services and infrastructure, and thus lead to major safety issues. Security risks in federated learning can originate through data or model poisoning (backdoor attacks), or when an adversary alters the data at inference time (evasion attack). Backdoor attacks pose one of the biggest challenges in federated learning as they are especially hard to detect. The challenge is increased by the secret nature of local training data, which makes it hard to analyse the correctness of the contribution of clients. Malicious clients can train models on poisoned data or directly manipulate model updates [3]. An adversary who is able to compromise the aggregator can perform attacks on the global model. Another threat comes from non-secure communication channels when an adversary is able to steal or maliciously modify shared model updates.

Privacy risks (confidentiality). Model parameters exchanged during federated learning represent an abstraction of the training data. Adversaries might infer information about training data having access to the model. In smart cities, data generated by sensors and IoT devices often involves personal privacy, and this is thus a great concern. It is thus important to mitigate potential leaks of this data through the machine learning process. Federated learning with the increased exchange of models might, however, increase the attack surface. Adversaries can perform different attacks on shared models in federated learning, e.g., model inversion, trying to recreate the original samples from the model, or membership inference, aiming to infer the membership of some

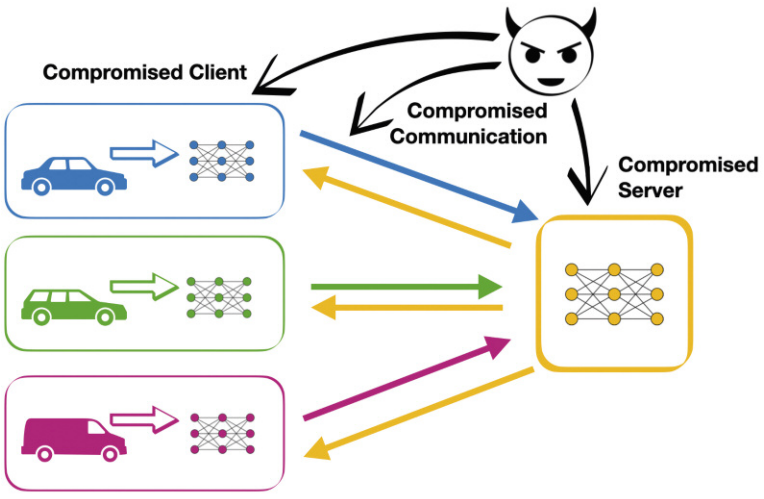


*Figure 1: Federated learning architecture and attack vectors. An adversary who is able to compromise clients, a server or communication channels can threaten the security and privacy of the system.*

particular instance in the training set of a target model [2]. Adversaries can be e.g., a compromised or malicious aggregator, or someone stealing models during client-server communication.

Approaches for mitigating security and privacy risks in federated learning often still lag behind attacks, but are increasingly in the focus of research activities.

Regarding privacy risks, several approaches can be employed. Differential privacy (DP) aims to bring uncertainty into the model outputs to hide personal contributions to the model; clients can add noise to shared model parameters or train a differentially private machine learning algorithm. The main downside of this approach remains that noise degrades models performance, thus there is a trade-off between privacy and utility.

Secure Multi-Party Computation (SMPC) provides a cryptographic protocol that allows joint computation of a function while keeping its inputs private. In federated learning, this can replace a central aggregator. However, SMPC poses high computational costs, therefore limiting the scalability of federated learning.

Homomorphic Encryption (HE) allows mathematical operations to be performed on encrypted data. Clients can encrypt their model parameters, and the coordinator could aggregate them but not understand them. Like SMPC, HE greatly increases computational costs.

Detecting attacks on the integrity and availability of the machine learning process is even more difficult. Defences like anomaly detection and robust aggregation aim to discover potentially harmful models and eliminate their malicious influence on the global model. Yet they fail to detect targeted backdoor attacks, as poisoned models look and behave similarly to models that were trained without backdoor [3].

There has been a dramatic increase in interest in federated learning in recent years. Many companies, including Apple and Google, are already using federated learning for their services. Interest in this technology is especially high in medical applications and smart cities, where personal data is processed, and data privacy is a major concern. However, there are still challenges to address in federated learning. Mitigation of security and privacy risk is especially important for building trust in the technology. Further investigation of defence mechanisms is therefore critical for the successful application of federated learning.

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

**References:**

[1] P. Kairouz, H. Brendan McMahan, et al.: "Advances and Open Problems in Federated Learning", Foundations and Trends in Machine Learning: Vol. 14: No. 1–2, pp 1-210, 2021.

[2] A. Pustozerova and R. Mayer: "Information leaks in federated learning", in proc. of the Workshop on Decentralized IoT Systems and Security (DISS), 2020.

[3] N. Bouacida and P. Mohapatra: "Vulnerabilities in Federated Learning", in IEEE Access, vol. 9, pp. 63229-63249, 2021.

**Please contact:**

Anastasia Pustozerova
SBA Research, Austria
apustozerova@sba-research.org