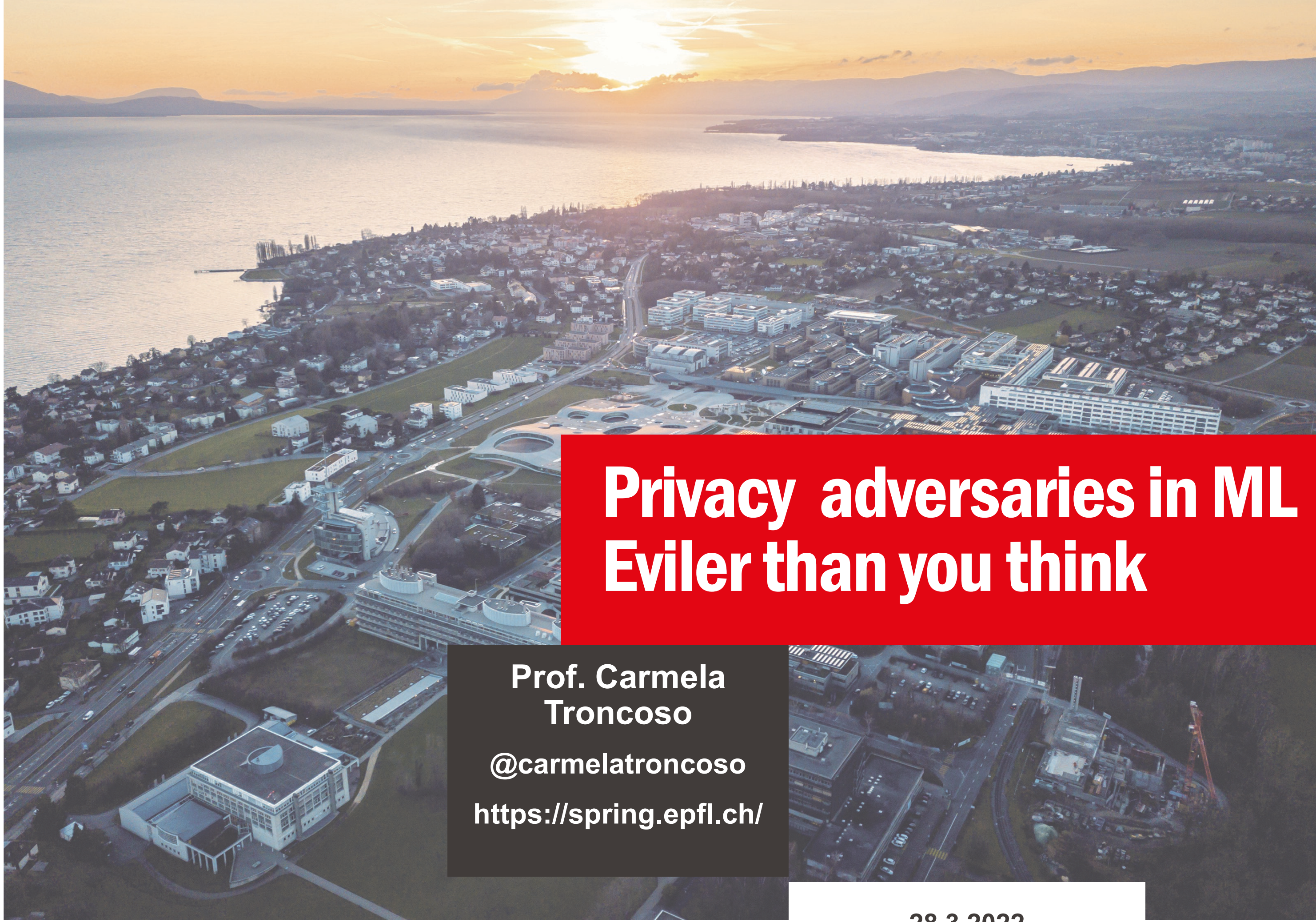
EPFL



Privacy adversaries in ML
Eviler than you think

Prof. Carmela Troncoso

@carmelatroncoso

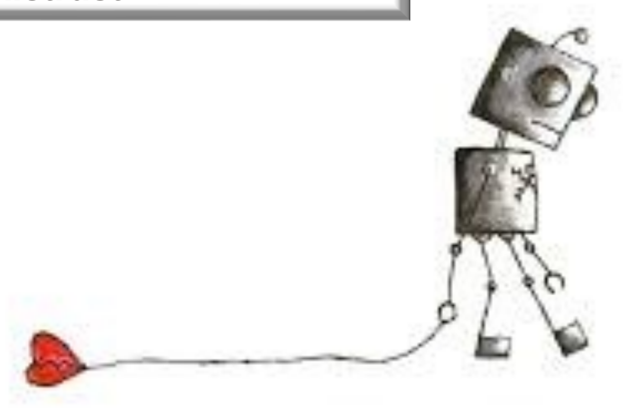
https://spring.epfl.ch/

28.3.2022

École polytechnique fédérale de Lausanne

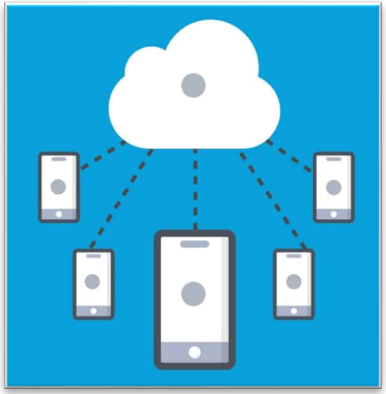
Privacy threatens Machine Learning

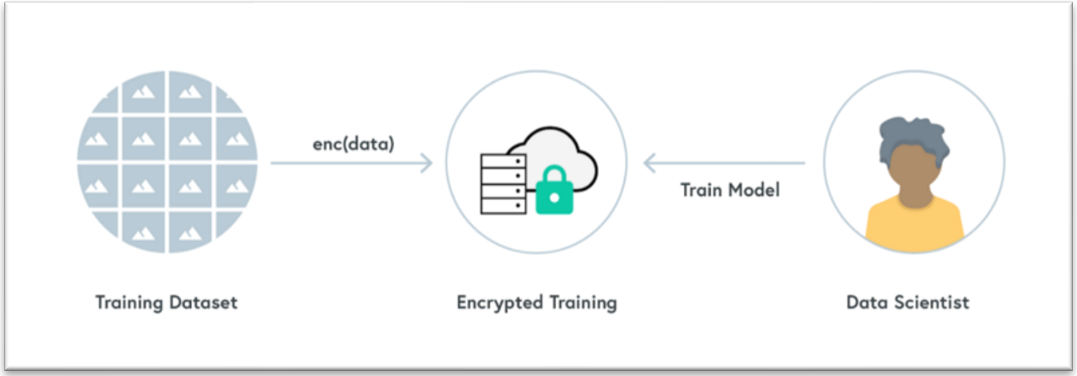
Invasive large-scale data collection results in users' mistrust

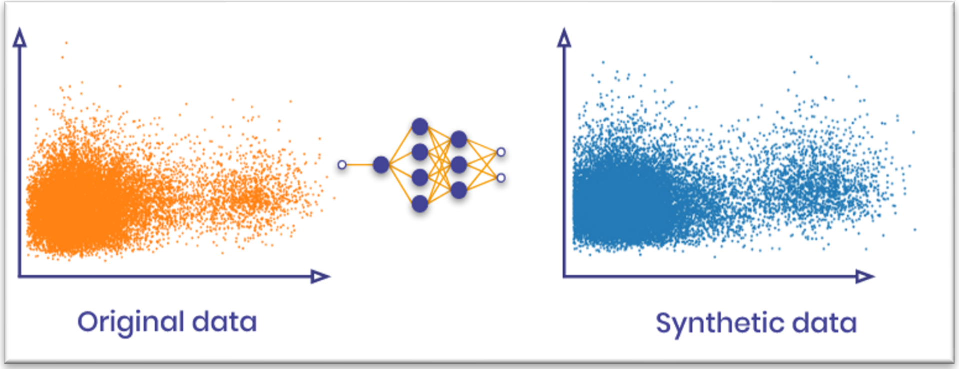


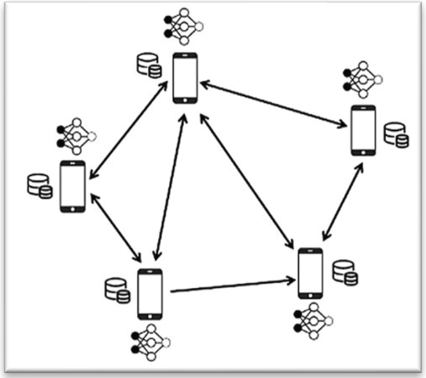
Regulations impose restrictions on data collection and processing

Privacy-preserving machine learning!









The problem…

“Honest-but-curious” adversary

“Non-strategic” adversary

The problem… and its consequences

WHEN THE CURIOUS ABANDON HONESTY: FEDERATED LEARNING IS NOT PRIVATE

Franziska Boenisch*
Fraunhofer AISEC
franziska.boenisch@aisec.fraunhofer.de

Adam Dziedzic†
University of Toronto and Vector Institute
adam.dziedzic@utoronto.ca

Ali Shahin Shamsabadi†
Vector Institute and The Alan Turing Institute
a.shahinshamsabadi@turing.ac.uk

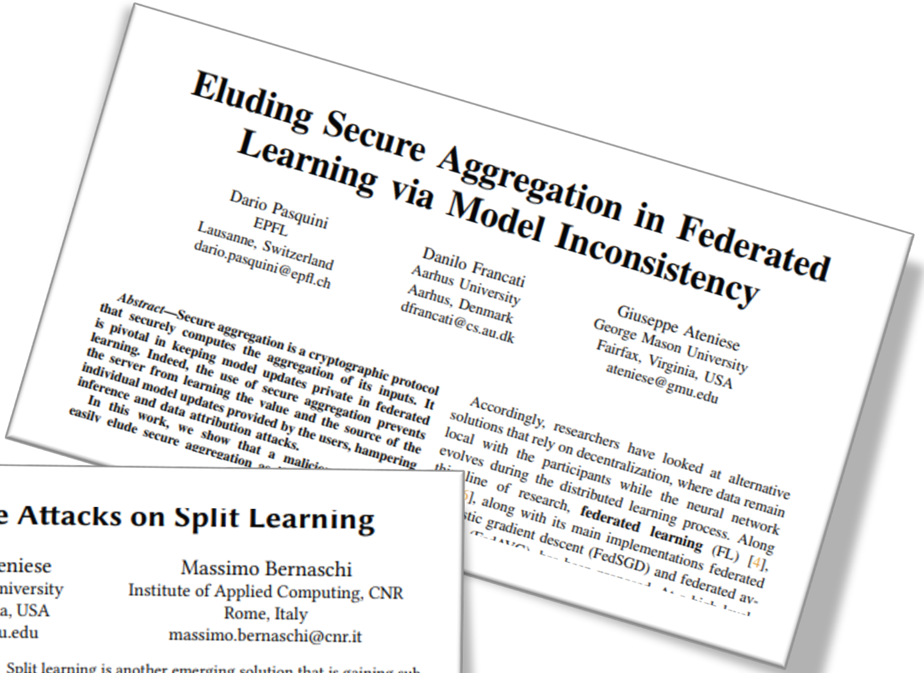
Ilia Shumailov†
Vector Institute
ilia.shumailov@cl.cam.ac.uk

Roei Schuster†
Vector Institute
roei@vectorinstitute.ai

Nicolas Papernot
University of Toronto and Vector Institute
…las.papernot@utoronto.ca

ABSTRACT

1 Introduction



Eluding Secure Aggregation in Federated Learning via Model Inconsistency

Dario Pasquini
EPFL
Lausanne, Switzerland
dario.pasquini@epfl.ch

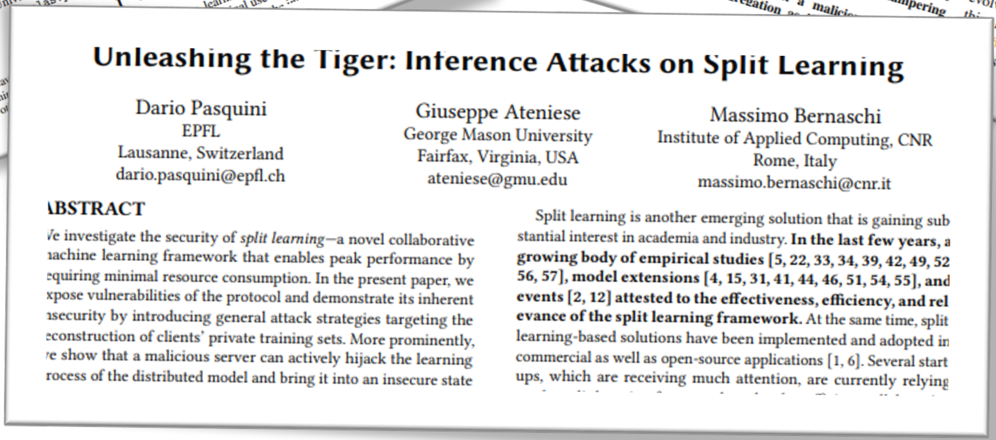
Danilo Francati
Aarhus University
Aarhus, Denmark
dfrancati@cs.au.dk

Giuseppe Ateniese
George Mason University
Fairfax, Virginia, USA
ateniese@gmu.edu

Abstract—**Secure aggregation is a cryptographic protocol that securely computes the aggregation of its inputs. It is pivotal in keeping model updates private in federated learning. Indeed, the use of secure aggregation prevents the server from learning the value and the source of the individual model updates provided by the users, hampering inference and data attribution attacks.**

In this work, we show that a malici… easily elude secure aggregation…

Accordingly, researchers have looked at alternative solutions that rely on decentralization, where data remain local with the participants while the neural network evolves during the distributed learning process. Along this line of research, **federated learning** (FL) [4], along with its main implementations federated av… stic gradient descent (FedSGD) and federated av-



Unleashing the Tiger: Inference Attacks on Split Learning

Dario Pasquini
EPFL
Lausanne, Switzerland
dario.pasquini@epfl.ch

Giuseppe Ateniese
George Mason University
Fairfax, Virginia, USA
ateniese@gmu.edu

Massimo Bernaschi
Institute of Applied Computing, CNR
Rome, Italy
massimo.bernaschi@cnr.it

ABSTRACT

…e investigate the security of *split learning*—a novel collaborative …achine learning framework that enables peak performance by …equiring minimal resource consumption. In the present paper, we …xpose vulnerabilities of the protocol and demonstrate its inherent …security by introducing general attack strategies targeting the …econstruction of clients' private training sets. More prominently, …e show that a malicious server can actively hijack the learning …rocess of the distributed model and bring it into an insecure state

Split learning is another emerging solution that is gaining substantial interest in academia and industry. **In the last few years, a growing body of empirical studies [5, 22, 33, 34, 39, 42, 49, 52, 56, 57], model extensions [4, 15, 31, 41, 44, 46, 51, 54, 55], and events [2, 12] attested to the effectiveness, efficiency, and relevance of the split learning framework.** At the same time, split learning-based solutions have been implemented and adopted in commercial as well as open-source applications [1, 6]. Several startups, which are receiving much attention, are currently relying

The problem… and its consequences

Synthetic Data – Anonymisation Groundhog Day

Theresa Stadler
EPFL

Bristena Oprisanu
UCL

Carmela Troncoso
EPFL

Abstract

Synthetic data has been advertised as a silver-bullet solution to privacy-preserving data publishing that addresses the shortcomings of traditional anonymisation techniques. The promise is that synthetic data drawn from generative models preserves the statistical properties of the original dataset but, at the same time, provides perfect protection against privacy attacks. In this work, we present the first quantitative evaluation of the privacy gain of synthetic data publishing and compare it to that of previous anonymisation techniques.

the board [11, 13, 14, 42, 44, 47, 58, 59]. A large number of publications, case studies, and real-world examples demonstrate that high-dimensional, sparse datasets are inherently vulnerable to privacy attacks. The repeated failures to protect the privacy of microdata releases reflect a fundamental trade-off: information-rich datasets that are valuable for statistical analysis also always contain enough information to conduct privacy attacks [45].

In this landscape, practitioners and researchers see in synthetic data a promising approach to open data sharing that addresses the privacy issues of previous anonymisation at-

ve show that a malicious server can actively hijack the learning rocess of the distributed model and bring it into an insecure state

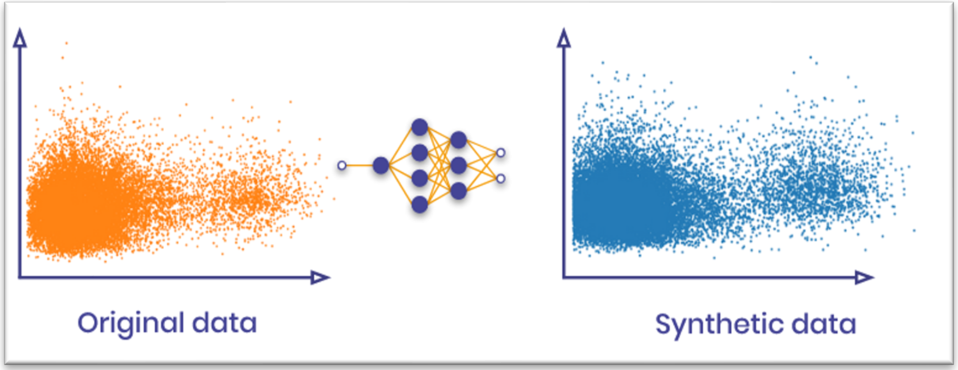
commercial as well as open-source applications [1, 6]. Several start ups, which are receiving much attention, are currently relying

Federated
tency

Ateniese
University
ia, USA
u.edu

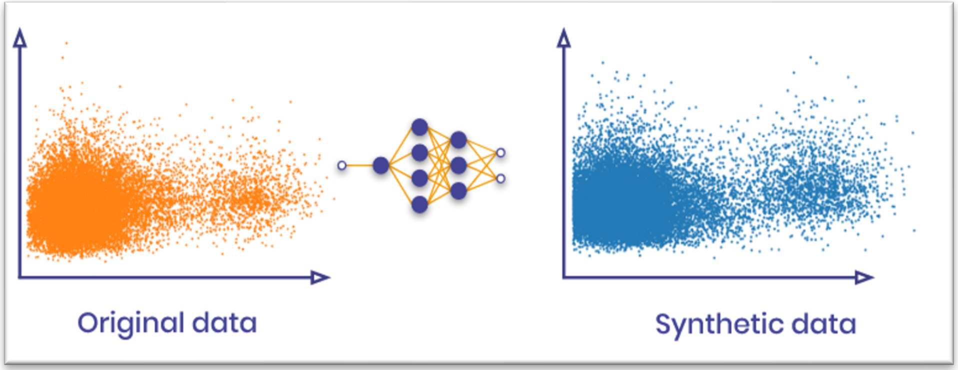
ked at alternative
where data remain
neural network
process. Along
ing (FL) [4],
ions federated
federated av-

Synthetic data is not a privacy-preserving mechanism



Synthetic data is private because there is no one-to-one mapping

Synthetic data is not a privacy-preserving mechanism

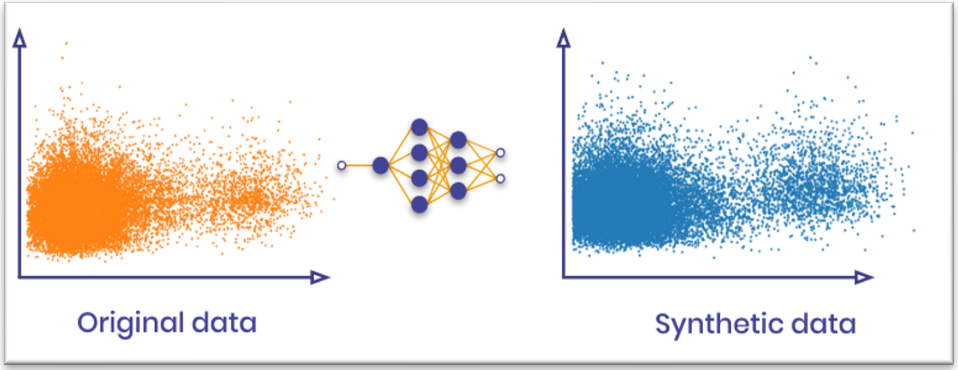


Synthetic data is private because there is no one-to-one mapping

Oh no! I can't make inferences anymore

Naïve adversary

Synthetic data is not a privacy-preserving mechanism



Synthetic data is private because there is no one-to-one mapping

Strategic adversary

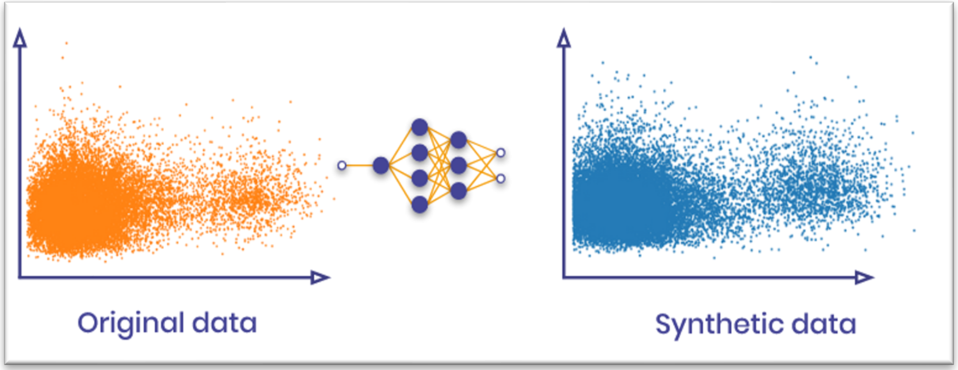
Oh no! I can't make inferences anymore

Naïve adversary

The information is preserved, same attacks are possible

Carmela Troncoso

Synthetic data is not a privacy-preserving mechanism



Synthetic data is private because there is no one-to-one mapping AND we add differential privacy

Strategic adversary

The information is still preserved because of implementation errors (recurrent across implementations as they increase utility)

Same attacks are possible

Food for thought

- Privacy adversaries must be as **evil** and **clever** as you can think
  - They are not honest: they will not follow protocol
  - They are strategic: they know the defense and will undermine it
  - … otherwise is not privacy, it is regulatory compliance

- Synthetic data is no silver bullet
  - If utility is preserved, so is information that enables inference attacks
  - If there is protection, it is not uniform for everyone and it is not predictable

- Empirical privacy evaluations are needed
  - Theory is hard in practice – always double check!