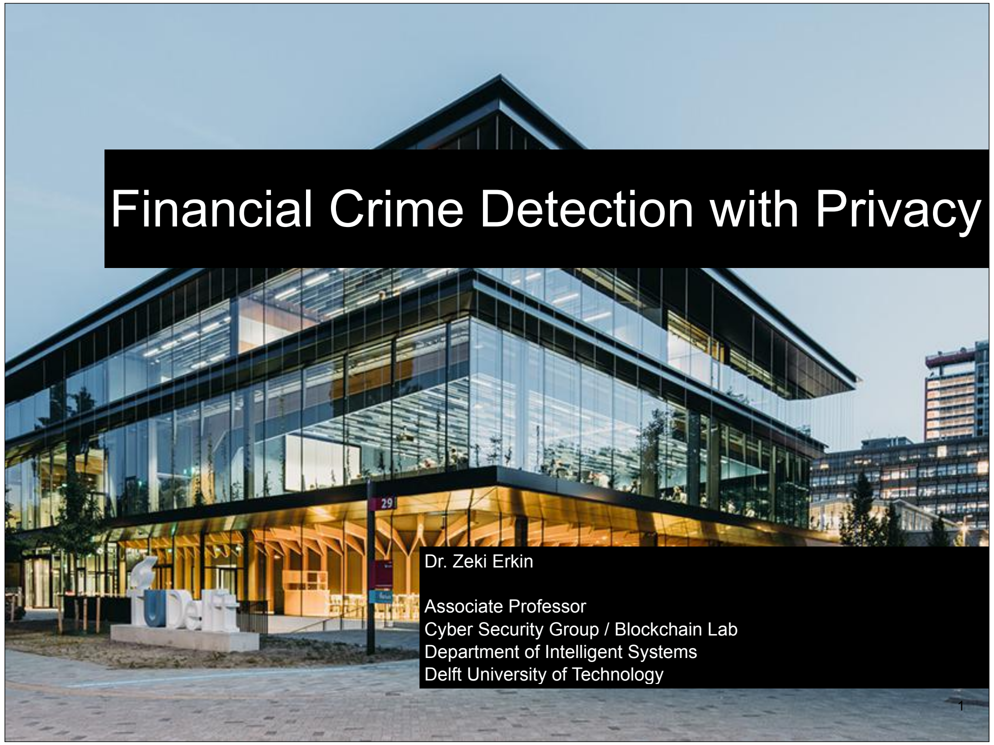


Financial Crime Detection with Privacy

Dr. Zeki Erkin

Associate Professor
Cyber Security Group / Blockchain Lab
Department of Intelligent Systems
Delft University of Technology

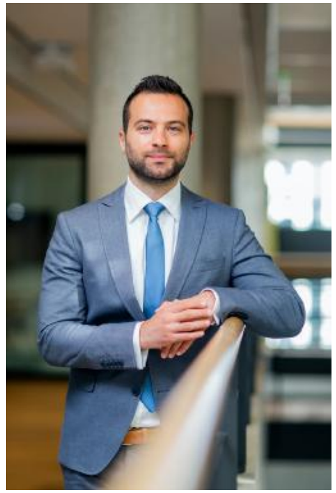
About me

- Associate Professor
- Cyber Security Group, Delft University of Technology

- **Research Interest**
  - Secure Information Sharing and Intelligence
  - Anonymisation
  - Decentralised Systems (DLT)

- **Teaching**
  - Security and Cryptography (MSc)
  - Privacy Enhancing Technologies (MSc)
  - Blockchain Engineering (MSc)

- IEEE SPS Information Forensics and Security TC chair
- EiC for Eurasip Journal on Information Security, Springer OPEN
- ACCSS vice-chair




MARCH 31, 2023

At Summit for Democracy, the United States and the United Kingdom Announce Winners of Challenge to Drive Innovation in Privacy-enhancing Technologies That Reinforce Democratic Values

OSTP › NEWS & UPDATES › PRESS RELEASES

- Transforming financial crime prevention and
- Boosting pandemic response capabilities through privacy-preserving federated learning

TU Delft

PET Prize Challenges

“The winning solutions combined different PETs to allow the AI models to learn to make better predictions without exposing any sensitive data.”

- Drive innovation
- Deliver strong end-to-end privacy guarantees
- Develop a privacy-preserving solution

TU Delft

Financial Crime Prevention

- Money laundering, 2 Trillion $ per year
- Privacy-preserving federated learning solutions
  - To detect anomalous payments
  - A combination of input and output privacy
  - Synthetic datasets from SWIFT

Datasets

- **D1**: A synthetic dataset representing transaction data created by SWIFT, the global provider of secure financial messaging services

- **D2:** Synthetic customer / account metadata flags representative of data held by banks

4 Million rows across the two datasets

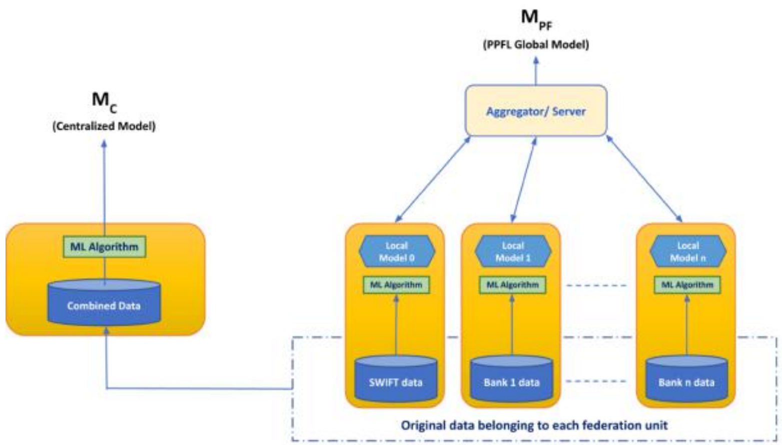
D1 Fields

MessageId	UETR	Sender	Receiver	OrderingAccount	BeneficiaryAccount	...
...	...	...	...	...	...	...
10	00012345-...	A	B	111	222	...
11	00012345-...	B	C	111	222	
12	00012345-...	C	D	111	222	
...	...	...	...	...	...	...

D2 Fields

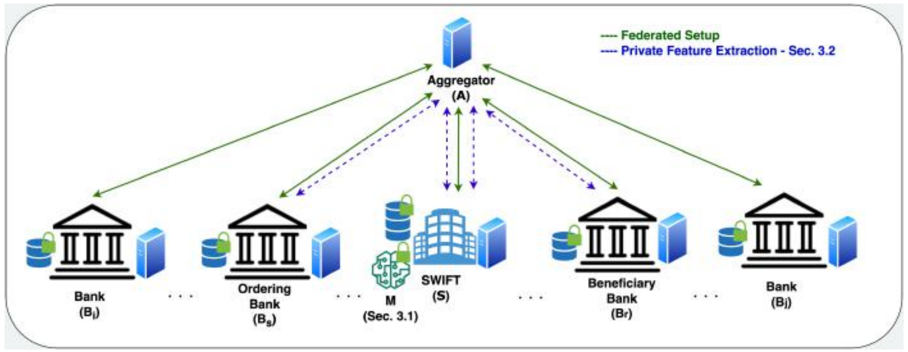
Street	Street address associated with the account
CountyCityZip	The remaining address details associated with the account
Flags	Enumerated data type indicating potential issues or special features that have been associated with an account. Flag definitions are: • 00 - No flags • 01 - Account closed • 03 - Account recently opened • 04 - Name mismatch • 05 - Account under monitoring • 06 - Account suspended • 07 - Account frozen • 08 - Non-transaction account • 09 - Beneficiary deceased • 10 - Invalid company ID • 11 - Invalid individual ID

Model



M_PF
(PPFL Global Model)
Aggregator/ Server
M_C
(Centralized Model)
ML Algorithm
Combined Data
Local Model 0
ML Algorithm
SWIFT data
Local Model 1
ML Algorithm
Bank 1 data
Local Model n
ML Algorithm
Bank n data
Original data belonging to each federation unit

Model



---- Federated Setup
---- Private Feature Extraction - Sec. 3.2
Aggregator
(A)
Bank
(Bi)
Ordering
Bank
(Bs)
M
(Sec. 3.1)
SWIFT
(S)
Beneficiary
Bank
(Br)
Bank
(Bj)

Evaluation Criteria

- The ability of the solution to deliver (and evidence) relevant privacy properties
- The accuracy of model MPF compared to MC
- The performance/computational cost of training MPF compared to MC
- The scalability, usability, and adaptability of the solution.

TU Delft

Timeline

TU Delft

PPML Huskies

Jelle Vos

- Martine De Cock, University of Washington Tacoma
- **Zekeriya Erkin, Delft University of Technology**
- Steven Golob, University of Washington Tacoma
- Dean Kelley, University of Washington Tacoma
- Ricardo Maia, University of Brasilia
- Anderson Nascimento, University of Washington Tacoma
- Sikha Pentyala, University of Washington Tacoma
- **C´elio Porsius Martins, Delft University of Technology**
- **Jelle Vos, Delft University of Technology**

Celio Porsius Martins

TUDelft

Money Laundering Detection

- Cross-silo federated architecture
- There are N Banks
- Communicating with a central entity S
- The Flower framework

- Train a model M
  - Input privacy: Encryption
  - Output privacy: Machine learning algorithm with Differential Privacy

- Custom tailored protocol
  - Elliptic curve El Gamal
  - Oblivious key-value stores (OKVS)
- Semi-honest security model

TUDelft

Privacy

- Input privacy: MPC
- Output privacy:
  - Model leaks information!
  - DP provides output privacy

Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pages 1322–1333, 2015.

Nicholas Carlini, Chang Liu, ́Ulfar Erlingsson, Jernej Kos, and Dawn Song. The secret sharer: Evaluating and testing unintended memorization in neural networks. In 28th USENIX Security Symposium, pages 267–284, 2019.

TUDelft

Our model

- SWIFT trains a local model (logistic regression)
- Training uses differential privacy to hide relation to the training set
- The outputs of the classifier therefore do not leak information about the training set
- The output is a probability that the transaction is fraudulent
- We always predict the transaction to be fraudulent if user’s data is **inconsistent**…

TUDelft

Our cryptographic protocol

- Performs a consistency check the sending and receiving users' data between SWIFT and a bank
  - Equivalent to two private set membership checks and an AND operation
  - The majority of the computation only has to be performed once on a bank's data
  - After that, queries only take ~a dozen elliptic curve multiplications

Experimental Results

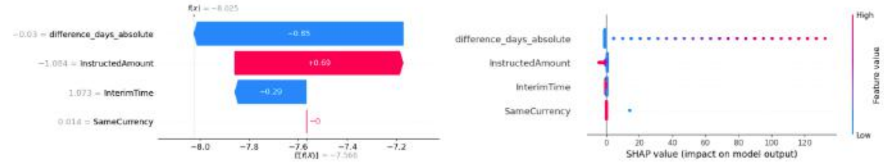


Figure 2: Visualization of SHAP values of an LR model trained with DP guarantees (see LR-DP in Sec. 5) to illustrate the effect of individual features on the model output.

Experimental Results

AUPRC:area under Precision-Recall (PR) curve

AUPRC	privacy	RF	LR	MLP	LR_{best}
with DP	$\epsilon = 0.5$	0.667	0.550	0.741	0.93
	$\epsilon = 1.0$	0.742	0.749	0.771	0.93
	$\epsilon = 5.0$	0.671	0.757	0.776	0.941
without DP	$\epsilon = \infty$	0.976	0.803	0.776	0.943

Random Forest, Linear Regression and Multilayer Percepton

	Time			Memory		Communication	
	Total	SWIFT	node	SWIFT	node	SWIFT	node
scenario 1 SWIFT + 2 nodes	1596s	1198s	228s	3.50GB	1.95GB	1052B	1584B
scenario 2 SWIFT + 4 nodes	1581s	1173s	234s	3.92GB	2.01GB	1200B	3168B
scenario 3 SWIFT + 9 nodes	2701s	2215s	243s	4.36GB	1.85GB	2236B	7128B

Desktop Intel i7 6700k at 4.2GHz, 64GB memory, and GTX1080 GPU

Official Results

	Entry	Method		C	N1	N2	N3
centralized	1	RF 8 features	OKVS 2 fields	0.8841			
	2	RF 4 features	OKVS 2 fields	0.9739			
	3	RF 4 features	OKVS 4 fields	0.9801			
federated	1	MLP with DP-SGD ($\epsilon = 5$) 4 features	OKVS 2 fields		0.8195	0.8235	0.8074
	2	LR with DP-SGD ($\epsilon = 5$) `bin_features`, `SameCurrency`	OKVS 4 fields		0.9494	0.9610	0.9477

- 8 features: `InstructedAmount`, `InterimTime`, `SettlementAmount`,`hour`,`sender_hour_freq`, `sender_currency_freq`,`sender_currency_amount_average`,`sender_receiver_freq`
- 4 features: `InstructedAmount`, `SameCurrency`, `InterimTime`, `difference_days_absolute`
- 2 fields: `Account`, `Name`
- 4 fields: `Account`, `Name`, `Street`, `CountryCityZIP`

And…

- PET is here!
  - Practical and scalable
- Team work was productive!
- But caution is needed …
  - Our solution is explainable
  - Not interpretable…

TUDelft

Why is interpretability important?

Dutch childcare benefits scandal

Article Talk

From Wikipedia, the free encyclopedia

This article needs to be **updated**. Please help update this article to reflect recent information. *(December 2021)*

This article **may be expanded with text translated from** the corresponding article i
2021) Click [show] for important translation instructions.

The **Dutch childcare benefits scandal** (Dutch: *kinderopvangtoeslagaffaire* or *toeslagenaffaire*, lit. '[childcare] benefits affair') is a political scandal in the Netherlands concerning false allegations of fraud made by the Tax and Customs Administration while attempting to regulate the distribution of childcare benefits.[1][2] Between 2005 and 2019, authorities wrongly accused an estimated 26,000 parents of making fraudulent benefit claims, requiring them to pay back the allowances they had received in their entirety.[1][3] In many cases, this sum amounted to tens of thousands of euros, driving families into severe financial hardship.[1][2]

The scandal was brought to public attention in September 2018. Investigators have subsequently described the working procedure of the Tax and Customs Administration as "discriminatory" and filled with "institutional bias".[4][5] On 15 January 2021, two months before the 2021 general election, the third Rutte cabinet resigned over the scandal following a parliamentary inquiry into the matter, which concluded that "fundamental principles of the rule of law" had been violated.[1][2][6]

TUDelft

US Winners

Final Winners:

Track A: Financial Crime Prevention

Scarlet Pets (Rutgers University)

PPML Huskies (University of Washington Tacoma, Delft University of Technology, University of Brasilia)

ILLIDAN Lab (Michigan State University, University of Calgary)

Demo Day

- May 22, London
- Free but required registration

Thank you!

PPMLHuskies

Place: 2nd in Track A: Financial Crime Prevention

Prize: $50,000

Team members: Martine De Cock, Anderson Nascimento, Sikha Pentyala, Steven Golob, Dean Kelley, Zekeriya Erkin, Jelle Vos, Célio Porsius Martins, Ricardo Maia

Our cryptographic protocol

1. Let a bank encode a hash of each user record into an oblivious key-value store
2. The OKVS returns an encryption of zero if the hash is contained in it
3. Query the OKVS of the sending bank and the receiving bank on the users' data, and sum up the ciphertexts homomorphically
4. Homomorphically multiply by a random value and collaboratively decrypt
5. Check if the resulting value is non-zero!