João P. Vilela

INESCTEC, CISUC & University of Porto, Portugal

Prediction of User Privacy Preferences in Mobile Devices via Federated Learning

(ERCIM Workshop on Privacy, Transparency, Sovereignty and Security, April 2023)

Joint work with: Alastair Beresford (UCambridge, UK), Ricardo Mendes (UCoimbra, PT), André Brandão (UPorto, PT) and Mariana Cunha (UPorto, PT)

snob-5g.com

cop-mode.dei.uc.pt

Research Group: Main Areas & Applications

- Security and Privacy:
  - Cryptography, provable security and formal verification
  - Network Security, privacy-enhancing technologies
  - Secure distributed systems, decentralized ID management
  - Trusted execution environments, secure edge streaming
- Networking:
  - Cloud and edge computing
  - Beyond 5G networks & services (slicing, orchestration, …)
  - Networks and systems management
  - Network virtualization and SDNs
- Application areas:
  - Smart Cities, Internet of Things, I4.0, Critical Infrastructures, Mobile Devices, ...

Research Group: Selected Recent Projects

- PRIVATEER - Privacy-first Security Enablers for 6G Networks (2023--) (HEurope SNS, Coord: SpaceHellas)
- ARCADIAN-IoT - Autonomous trust, security and privacy management framework for IoT (H2020, IPN Portugal)
- CyberSec4Europe - Cyber Security Network of Competence Centres for Europe (H2020, Univ. Frankfurt)
- DISCRETION - Disruptive SDN secure communications for European Defence (EDIDP, DEIMOS Engenharia)
- ATENA - Advanced Tools to assEss and mitigate the criticality of ICT compoNents and their dependencies over Critical InfrAstructures (H2020, Leonardo S.p.A.)
- POSEIDON - Protection and control of Secured Information by means of a privacy enhanced Dashboard (H2020, MEF)
- DARPA SIEVE - Securing Information for Encrypted Verification and evaluation (DARPA, SRI Subcontracted)
- Safe Cities - Building Urban Safety (P2020, Bosch)
- SafeCloud - Secure and Resilient Cloud Architecture (H2020, INESCTEC)
- SNOB5G - Scalable Network Backhauling for 5G (MIT-Portugal, Ubiwhere)
- AIDA - Adaptive, Intelligent and Distributed Assurance Platform (CMU-Portugal, Mobileum)
- COP-MODE - Context-aware Privacy protection for Mobile Devices (H2020 NGI-Trust)

The Problem of Privacy in Mobile Devices

Dozens of apps

x

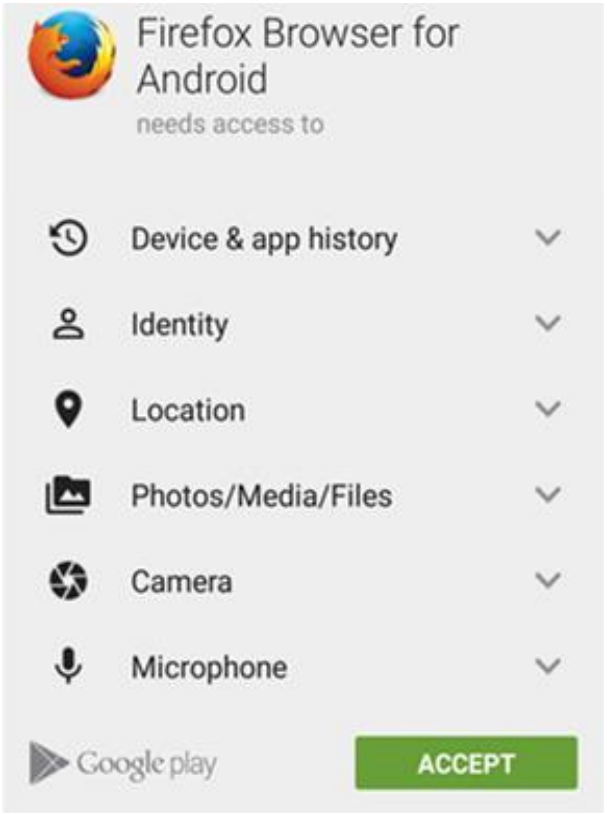
Multiple Configurations

=

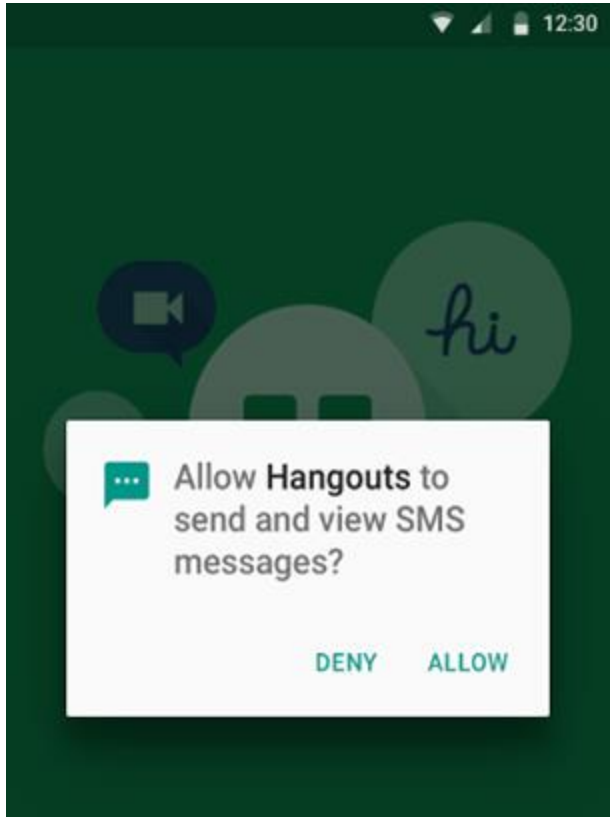
Privacy Loss

Privacy in Mobile Devices

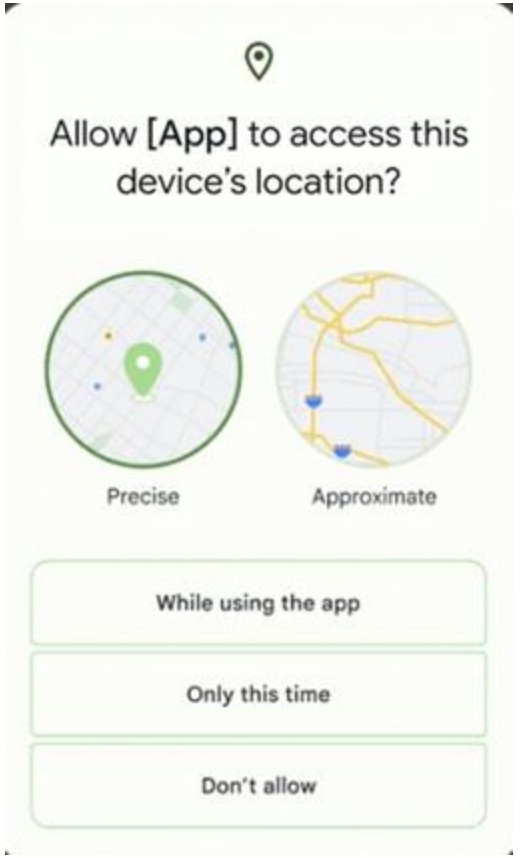
Install-time Permissions



Runtime Permissions (Oct 2015)

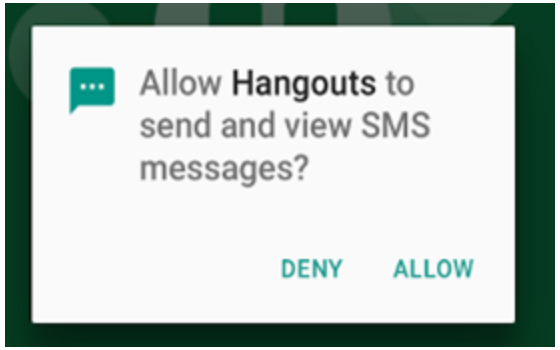


Latest Improvements



- Location Obfuscation
- "While using the app" (location, camera and microphone)
- "Only this time"
- Auto-reset when unused

Privacy in Mobile Devices



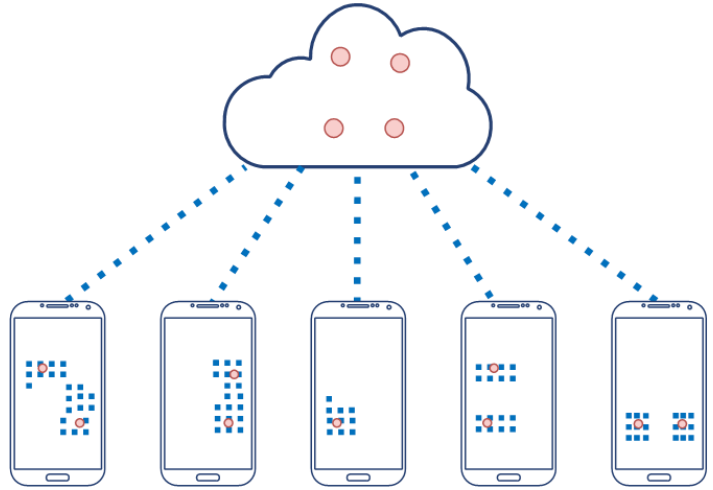
Runtime permissions allow:

- fine-grained permissions control
- to contextualize permission prompts by the needs of the app

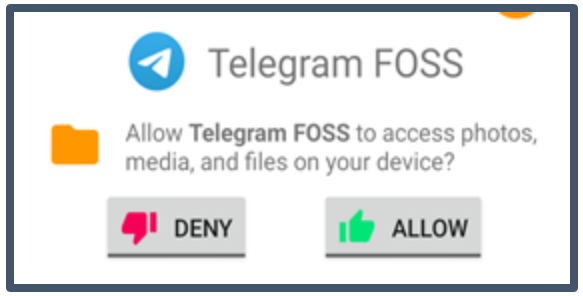
The problem: (hundreds of daily) **automatically accepted permissions**

- **Violate** contextual integrity (preferences of user within context)
- **Contradict** user expectations

Solution: Automated Privacy Decisions



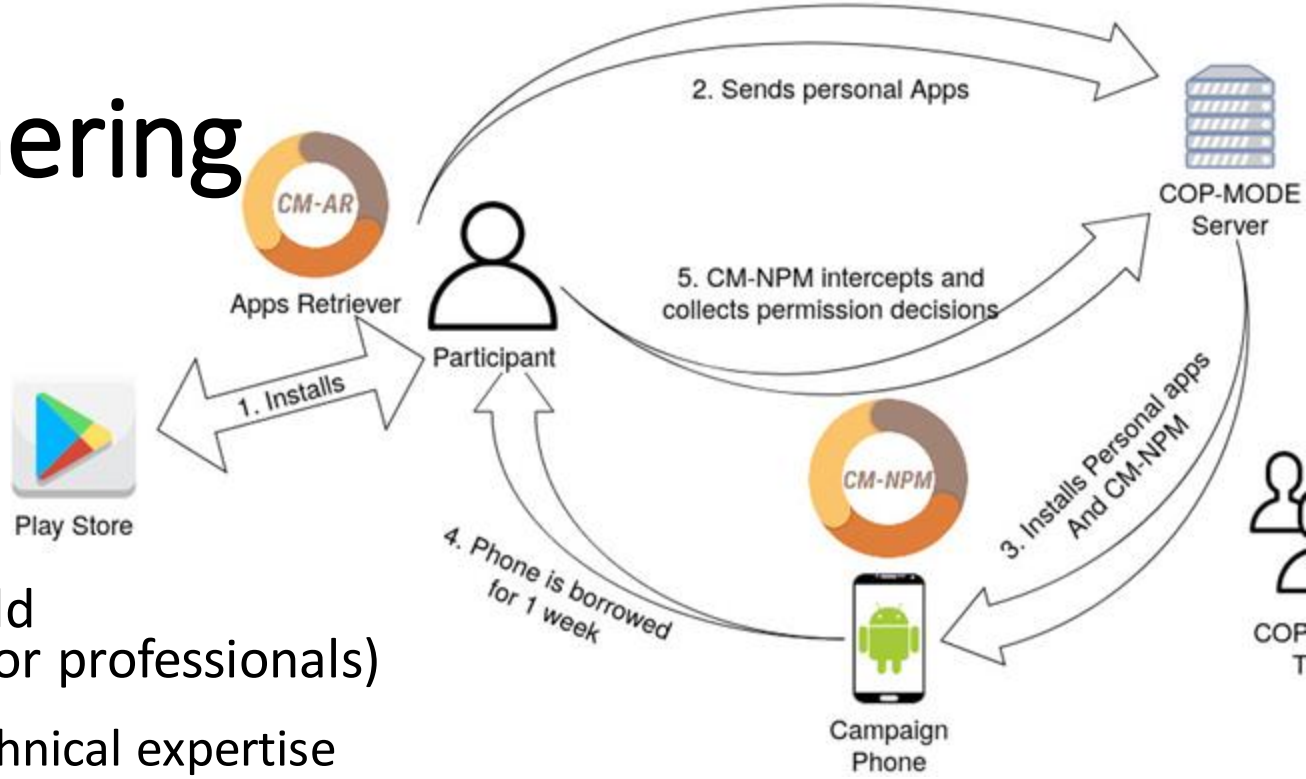
- Several devices with local info on:
  - Requesting Application
  - Permission
  - Grant Decision



- Prediction of Grant Decisions (Allow / Deny)
- According to Users' Preferences

For this, we need data!

COP-MODE: Data Gathering



93 participants

- 64.5% were students
- 71% were between 18 and 24 years old
- 57% with an IT background (studying or professionals)

→ Biased data towards young adults with technical expertise

Using our smartphones for 1 week+

Collected answers to 2M+ permission requests

65K+ manually answered requests

(~837/day, ~35/hour)

https://cop-mode.dei.uc.pt/campaigns

COP-MODE Data: Main Findings

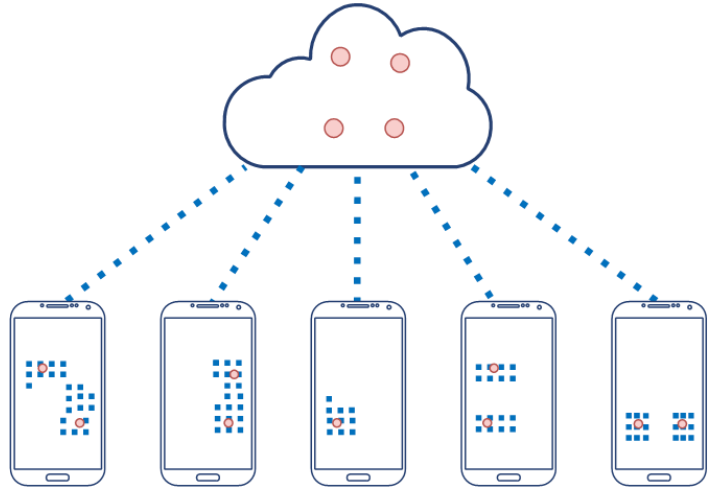
65K+ manually answered permission requests:

- Avg 836 requests/day/user, nearly 35/hour
- **Nearly 50% requests unexpected to users**
- **15% privacy violations**

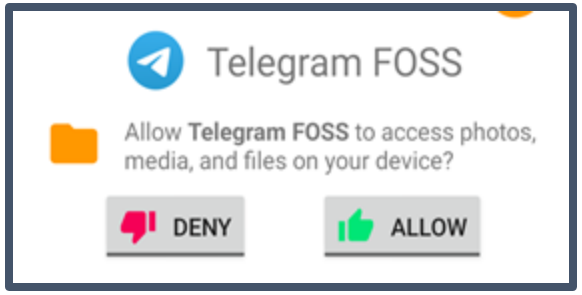
To catch 15% privacy violations ► answer 35 requests/hour

[Mendes et al., "Effect of User Expectation on Mobile App Privacy: A Field Study", PerCom'22]

Solution: Automated Privacy Decisions



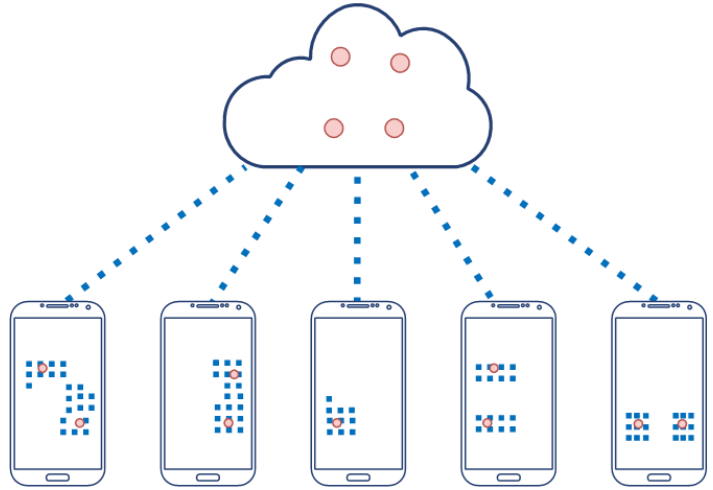
- Several devices with local info on:
  - Requesting Application
  - Permission
  - Grant Decision



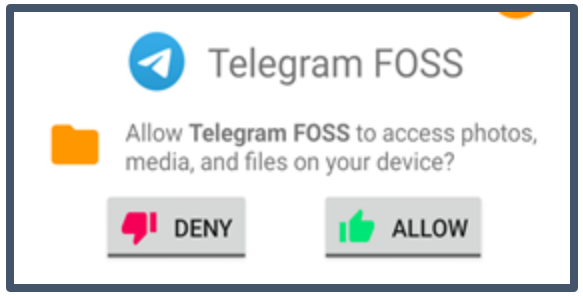
- Prediction of Grant Decisions (Allow / Deny)
- According to Users' Preferences

For this, we need data!

Solution: Automated Privacy Decisions



- Several devices with local info on:
  - Requesting Application
  - Permission
  - Grant Decision



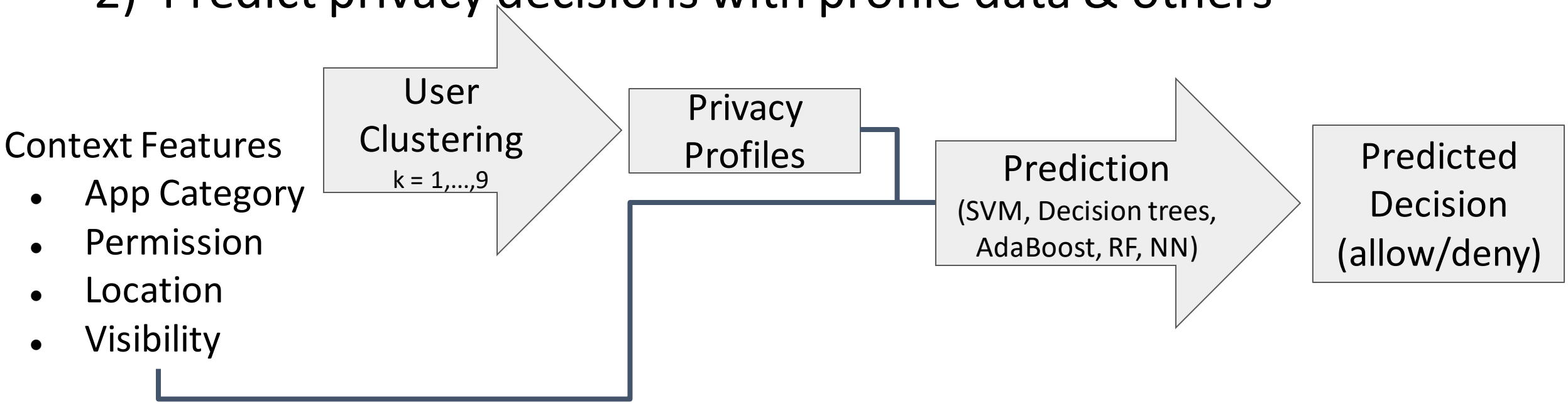
- Prediction of Grant Decisions (Allow / Deny)
- According to Users' Preferences

WITHOUT ACCESS/SHARING OF USER DATA

Prediction of Grant Decisions

A 2-stage approach:

1) Clustering users into profiles
2) Predict privacy decisions with profile data & others



Prediction of Grant Decisions with Privacy Guarantees

A 2-stage approach:

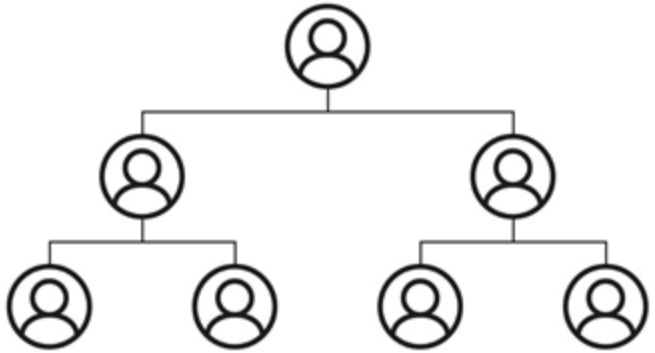
1) Clustering users into profiles
2) Predict privacy decisions with profile data & others

In a privacy-aware manner, i.e. without access to user data:

- Privacy-preserving clustering mechanisms
- Federated mechanisms for prediction of privacy decisions

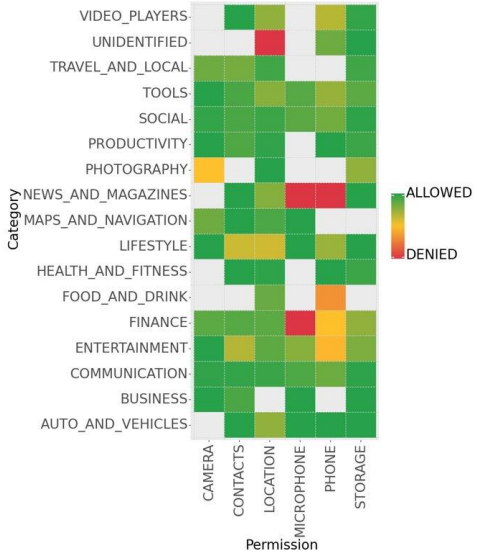
1) Secure Generation of Privacy Profiles

- Clustering of users into privacy profiles
  - (app category, permission, avg_grant_result)
- Profiles represent users' beliefs and expectations
- 2 approaches:
  - Distributed hierarchical clustering
  - Private k-means clustering

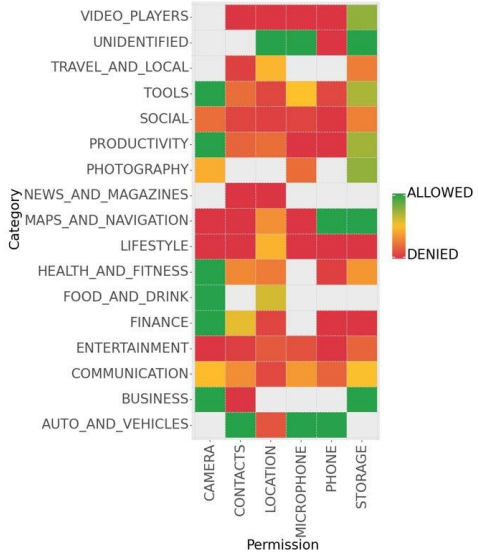


[Hamidi et al. PDP'18]

[Brandão et al. IDA'21]



Permissive User



Restrictive User

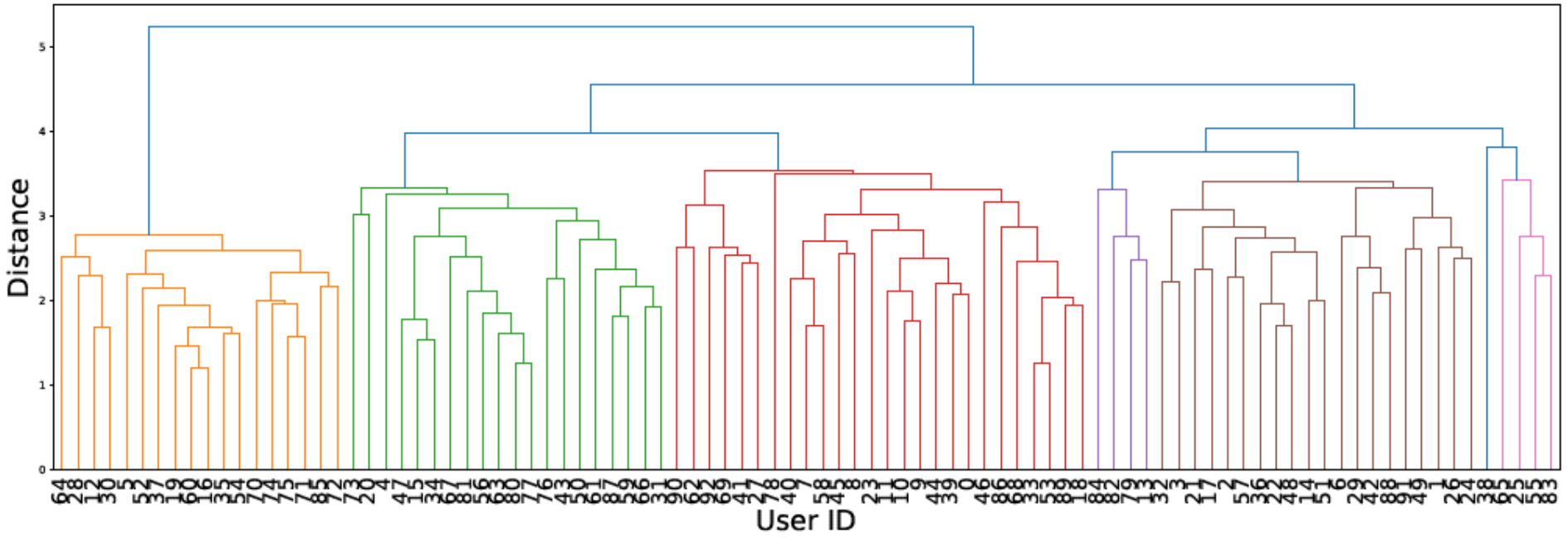
Distributed Hierarchical Clustering

[Hamidi et al. PDP'18]

- Per user (Category, Permission, Grant result):

App category:	EVENTS	EVENTS	...	AUTO_AND_VEHICLES	AUTO_AND_VEHICLES
Requested permission:	CALENDAR	CAMERA	...	PHONE	CONTACTS
Grant result:	0.9	0	...	0	0
	0.2	0.1	...	0.35	0.4

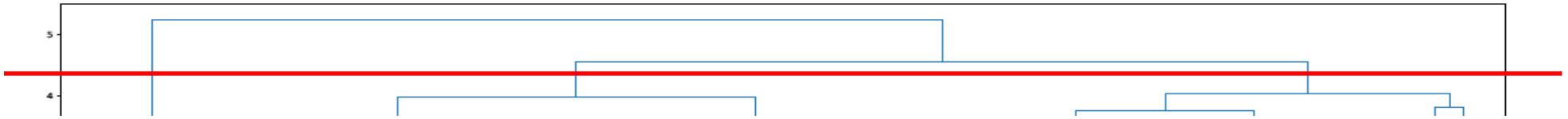
- Hierarchical clustering to divide users into profiles by creating a dendogram of distances and make a cut were appropriate



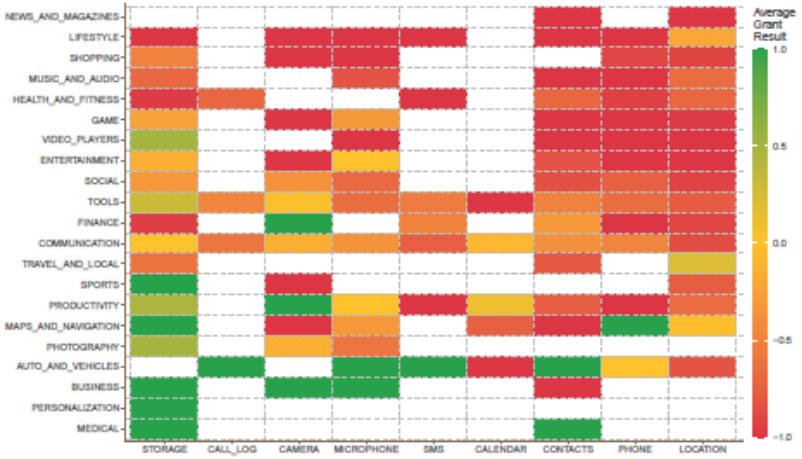
Distributed Hierarchical Clustering

[Hamidi et al. PDP'18]

- Making a cut at distance 4.3

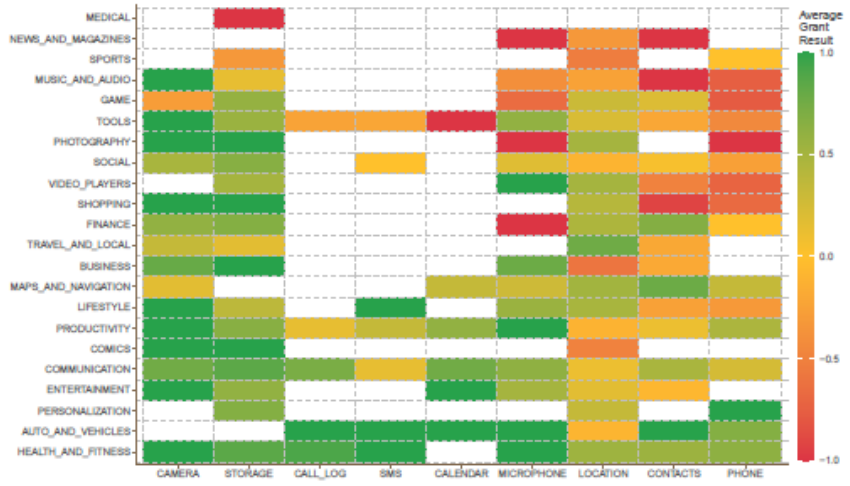


- Would result in 3 profiles as follows



(a) Profile 1 - the privacy conscious user.

(b) Profile 2 - permissive user.

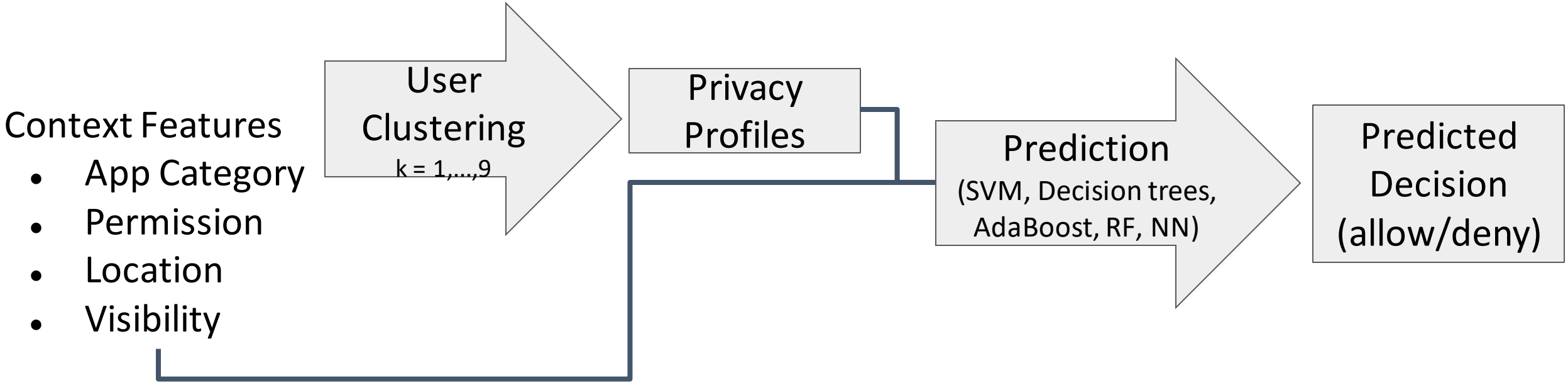


(c) Profile 3 - the middle ground user.

Prediction of Grant Decisions

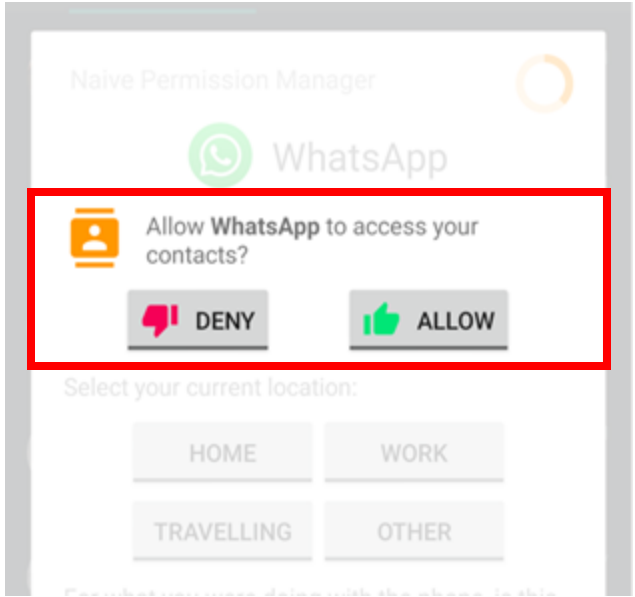
A 2-stage approach:

1) Clustering users into profiles
2) **Predict privacy decisions with profile data & others**



2) Grant Prediction with Federated Learning

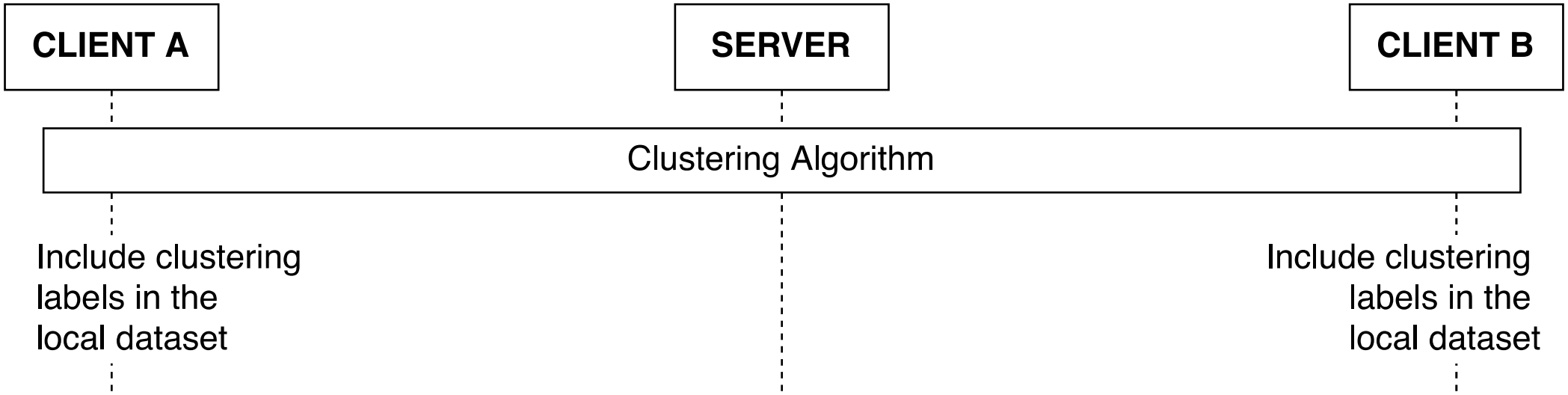
- Features:
  - **Profile (previous slide)**
  - app_category
  - isForeground
  - checkedPermissionGroup
  - isTopAppRequestingApp
  - checkedPermission
  - screenIsInteractive
  - hour
  - networkStatus
  - weekday
  - profile



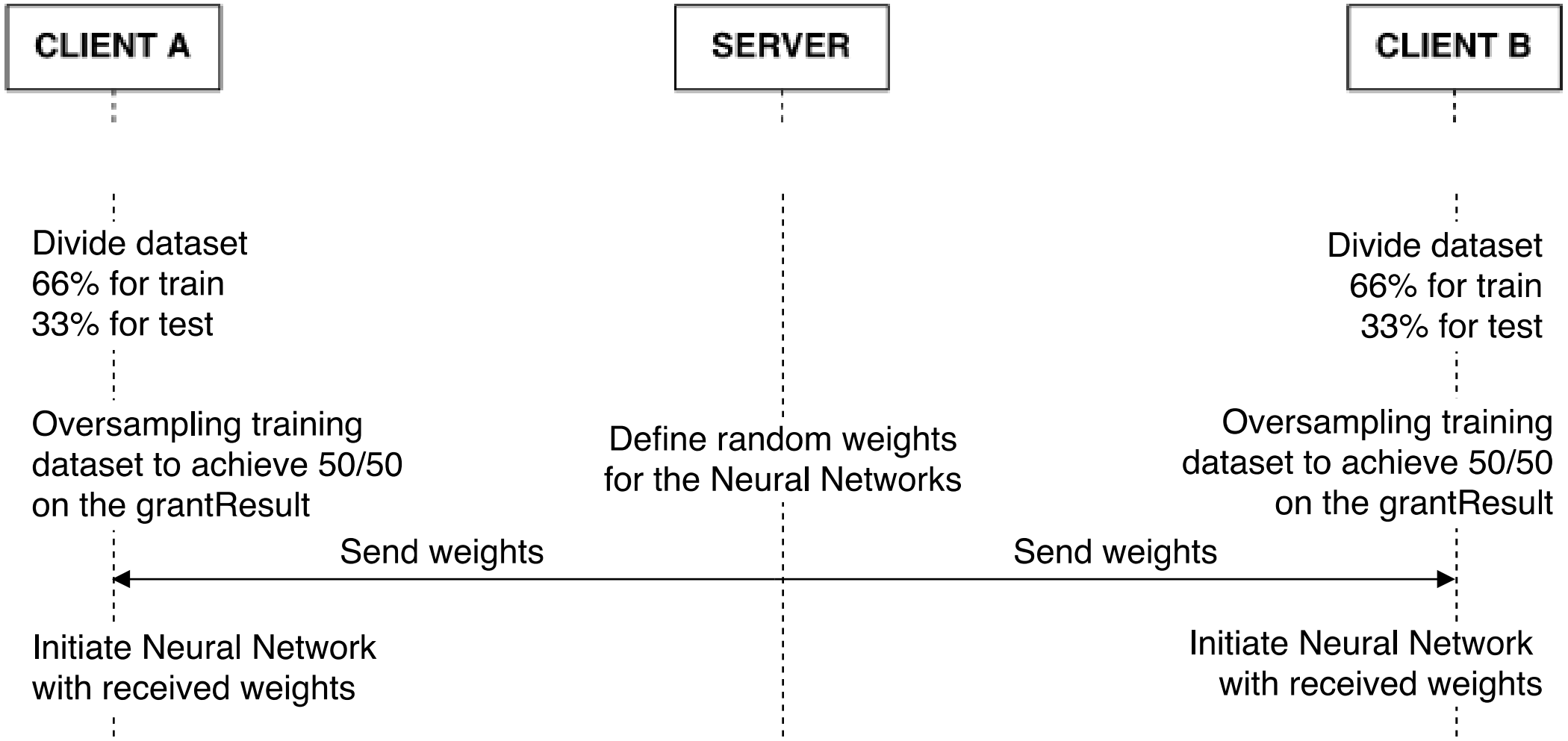
- Federated learning:
  1. Train neural network locally, on each smartphone, using only local data
  2. Share only the neural network weights (not the data) with a central server on each iteration

[Brandão et al., "Prediction of Mobile App Privacy Preferences with User Profiles via Federated Learning", CODASPY'22]

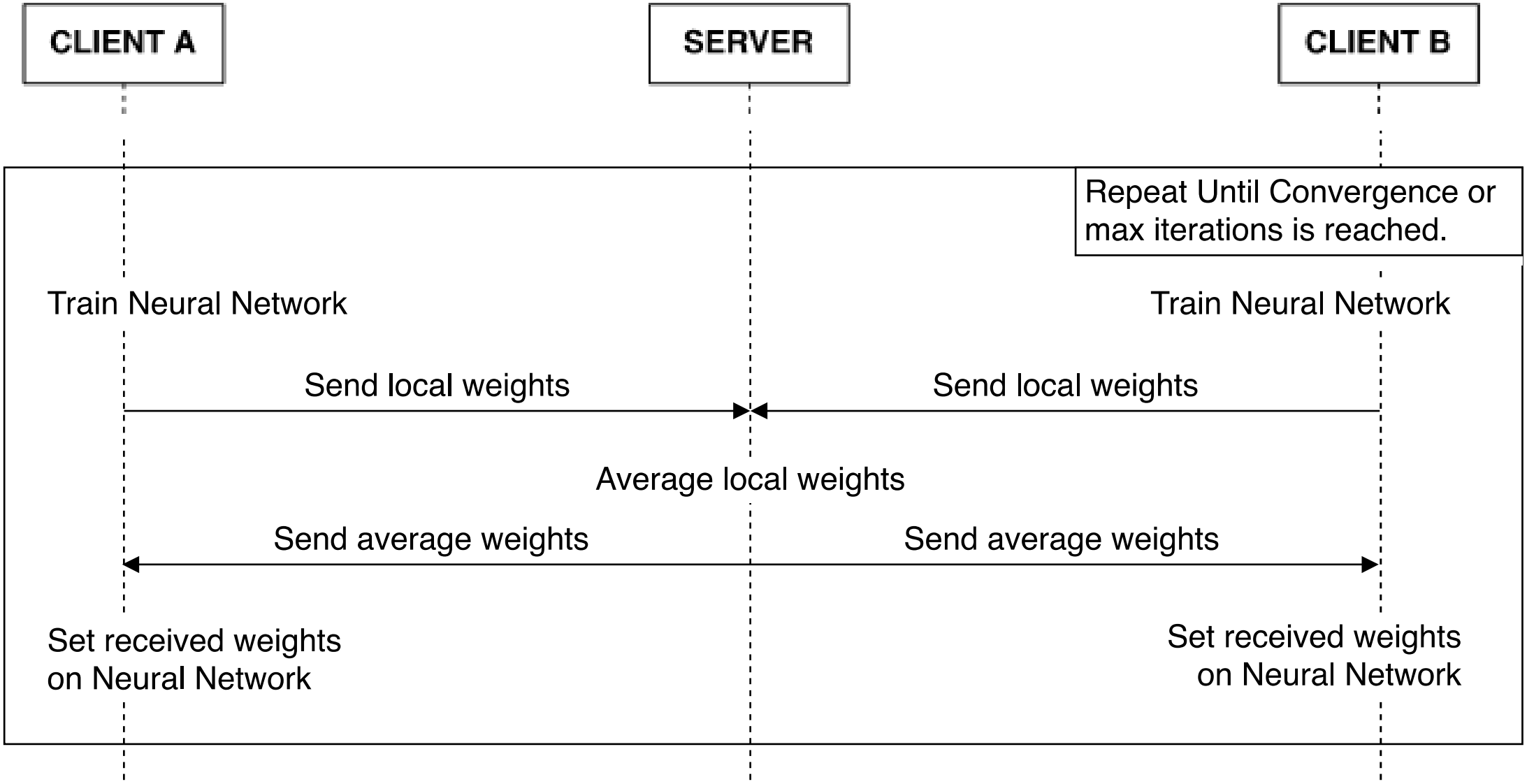
Federated Learning for Grant Prediction



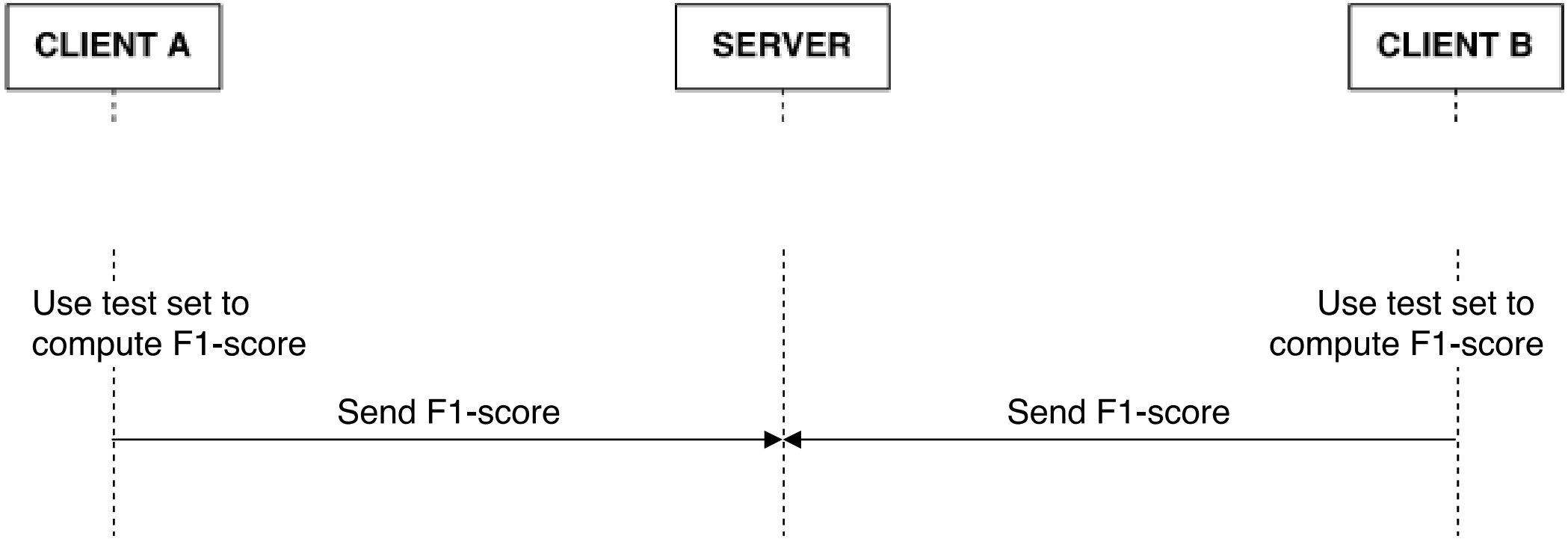
Federated Learning for Grant Prediction



Federated Learning for Grant Prediction



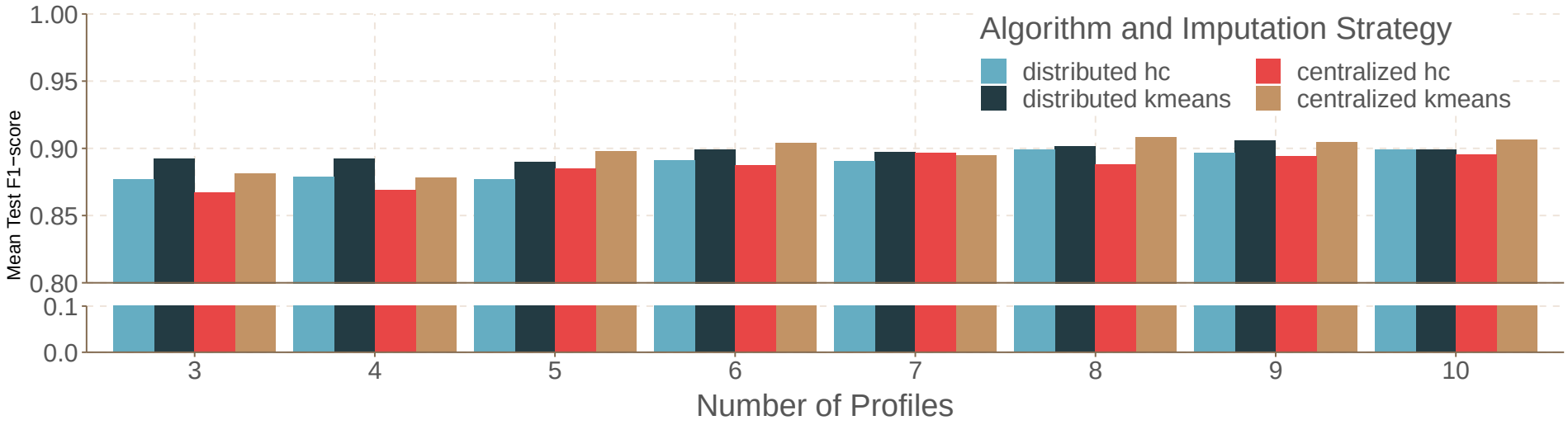
Federated Learning for Grant Prediction



Evaluation

Applied to our dataset of ~65K grant decisions from 93 users [Mendes et al. PerCom' 22]
http://cop-mode.dei.uc.pt/dataset

- Validation:
  - Grid search on the following parameters:
    - Clustering Algorithm.
    - Number of Clusters.
  - 5-fold cross validation with 80% of the dataset



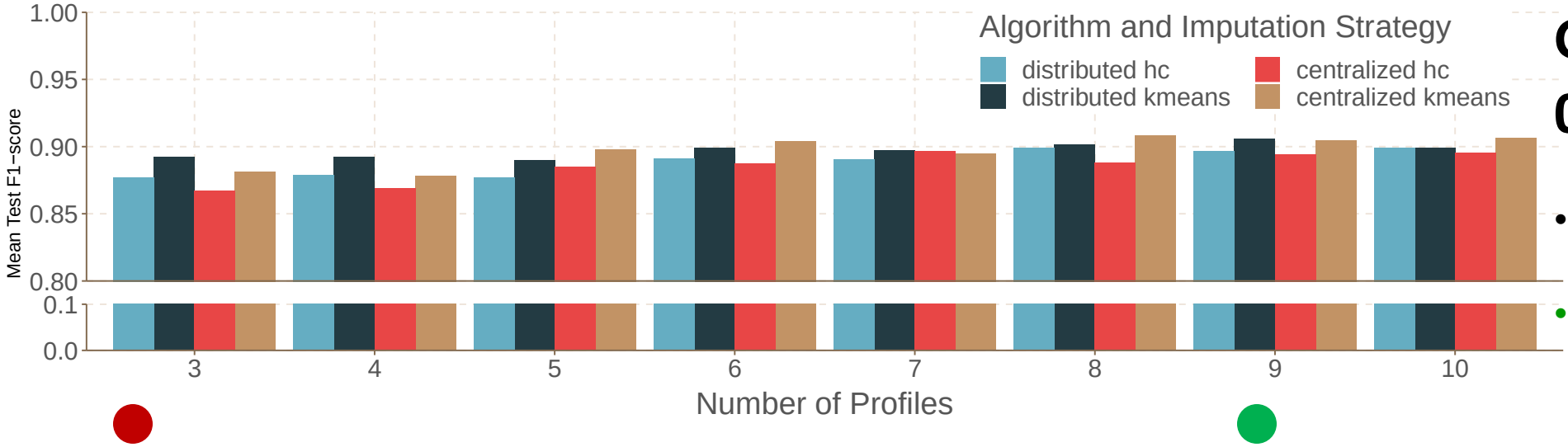
Evaluation

Applied to our dataset of ~65K grant decisions from 93 users [Mendes et al. PerCom' 22]
http://cop-mode.dei.uc.pt/dataset

- Validation:
  - Grid search on the following parameters:
    - Clustering Algorithm.
    - Number of Clusters.
  - 5-fold cross validation with 80% of the dataset

Best Mean F1-score of $\mathbf{0.91}$ with: Distributed k-Means ($k = 9$)

Worst Mean F1-score of $\mathbf{0.87}$ with: Distributed hc ($k = 3$)



Global F1-score: 0.91

- **Comparable to the centralized version**
- **Prediction and clustering in a privacy-preserving manner**

Conclusions and Future Work

- Privacy-preserving strategy to predict user's grant decisions
- Based on a 2-step approach:
  - Privacy-preserving clustering of users into profiles
  - Predict grant results through federated mechanisms
- Applied to a real world dataset of ~65K grant decisions from 93 users
- Maintain SoA prediction performance, **while preserving user privacy**
  - Reduces amount of privacy violations

- Future work:
  - Predicting User Expectation
  - DL + FL to replace the two-step process by a single one

Main References

1. Mendes, Brandão, Vilela, Beresford, "Effect of User Expectation on Mobile App Privacy: A Field Study", International Conference on Pervasive Computing and Communications (PerCom), 2022
2. Mendes, Cunha, Vilela, Beresford, "Enhancing User Privacy in Mobile Devices Through Prediction of Privacy Preferences", European Symposium on Research in Computer Security (ESORICS), 2022
3. Brandão, Mendes, Vilela, "Prediction of Mobile App Privacy Preferences with User Profiles via Federated Learning", ACM Conference on Data and Application Security and Privacy (CODASPY), 2022
4. Brandão, Mendes, and Vilela, "Efficient privacy preserving distributed K-means for non-IID data". In Advances in Intelligent Data Analysis XIX, 2021
5. Hamidi, Sheikhalishahi, and Martinelli, "A Secure Distributed Framework for Agglomerative Hierarchical Clustering Construction". In Euromicro International Conference on Parallel, Distributed and Network based Processing, 2018