Privacy preserving data analytics

Coimbra, 11 July, 2019

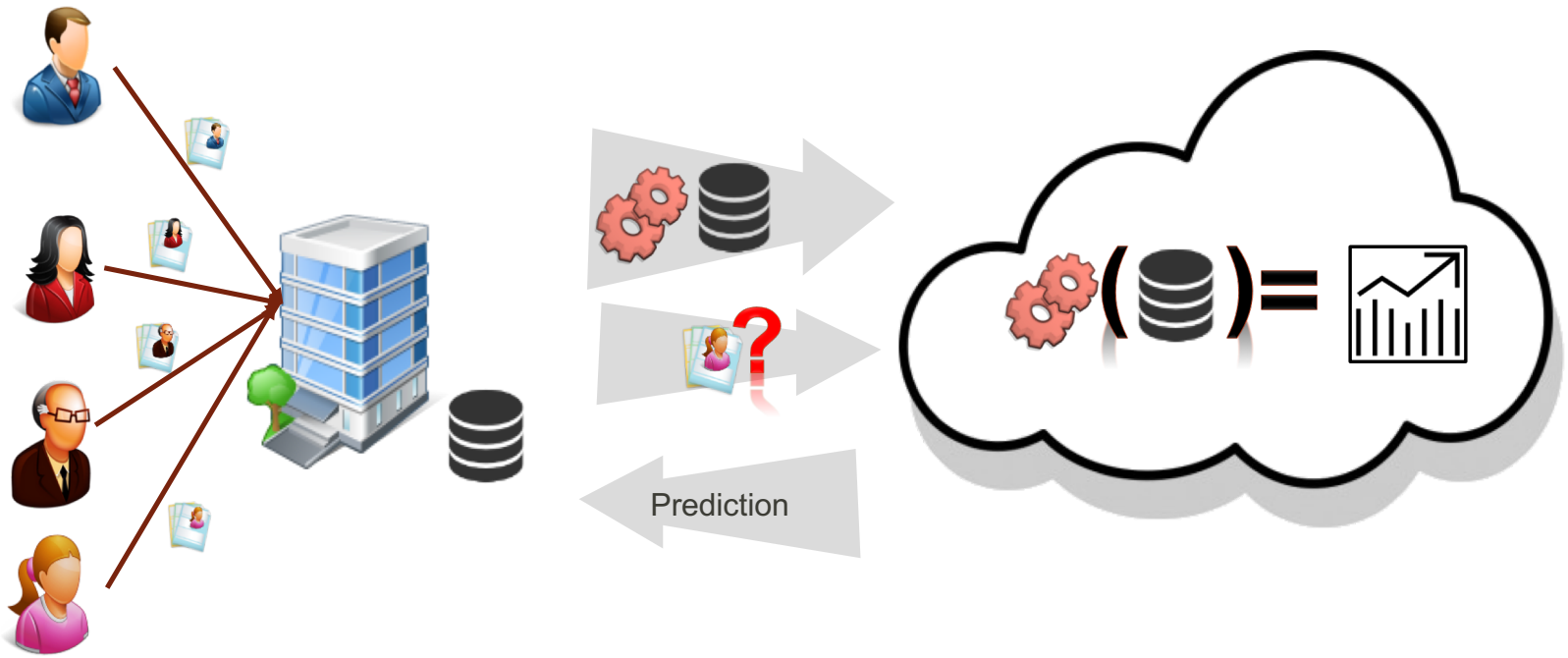
Melek Önen

Co-funded by the Horizon 2020 Framework Programme of the European Union

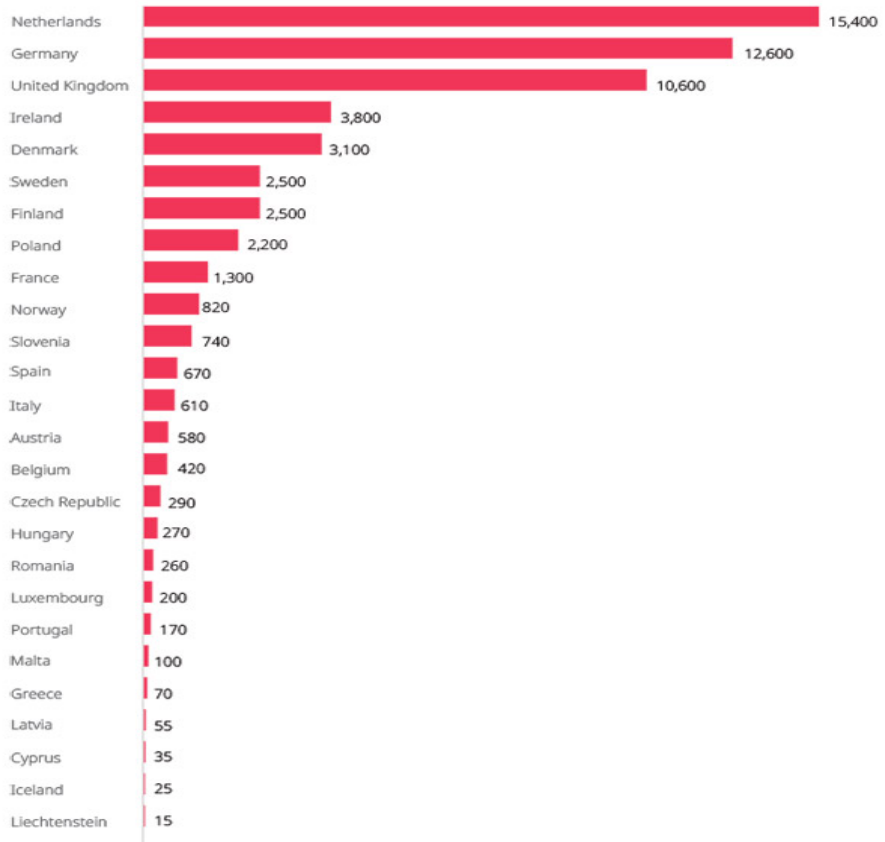
PAPAYA Project

Project title	**PAPAYA:** ***PlAtform for PrivAcY preserving data Analytics***
Call	**DS-08 Cybersecurity PPP: Privacy, Data Protection, Digital Identities**
Grant Agreement	**GA no: 786767**
Project Officer	**Mr. Nikolaos Panagiotarakis (H2020, REA)**
Duration	**36 months (1 May 2018 – 30 Apr 2021)**
Consortium	EURECOM, Atos, IBM, Karlstads Universitet, MEDIACLINICS Wearable Health Applications, orange
Project Coordinator	**EURECOM**

Data Analytics as a Service



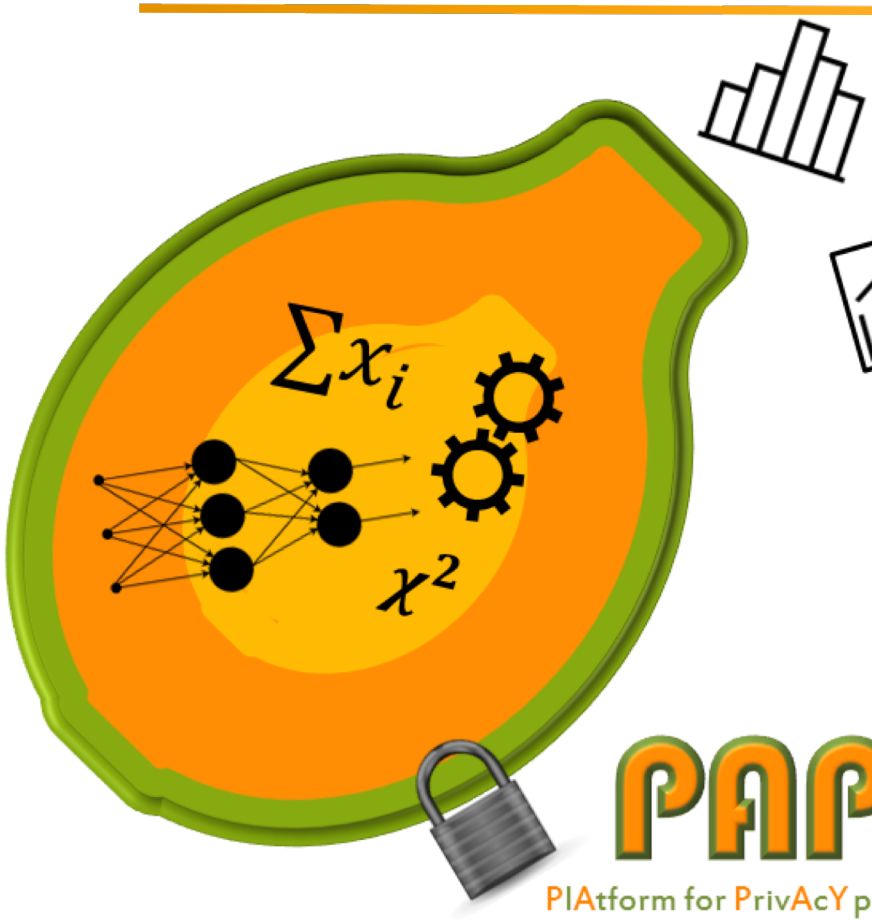
2019 Data Breaches in Europe



- # Data Breach Notifications
  - ~59K from May 2018 until Jan. 2019
- Average Cost in 2018
  - Global: 3.96M$, Per record: 148$
- Top 3 sectors
  - Health, Financial, Services
- Factors increasing cost
  - Extensive migration to cloud
  - Third party involvement
  - Compliance failures
- Factors decreasing cost
  - Extensive use of encryption
  - Use of security analytics

https://www.bankinfosecurity.com/gdpr-data-breach-reports-to-eu-exceed-59000-a-12006

PAPAYA

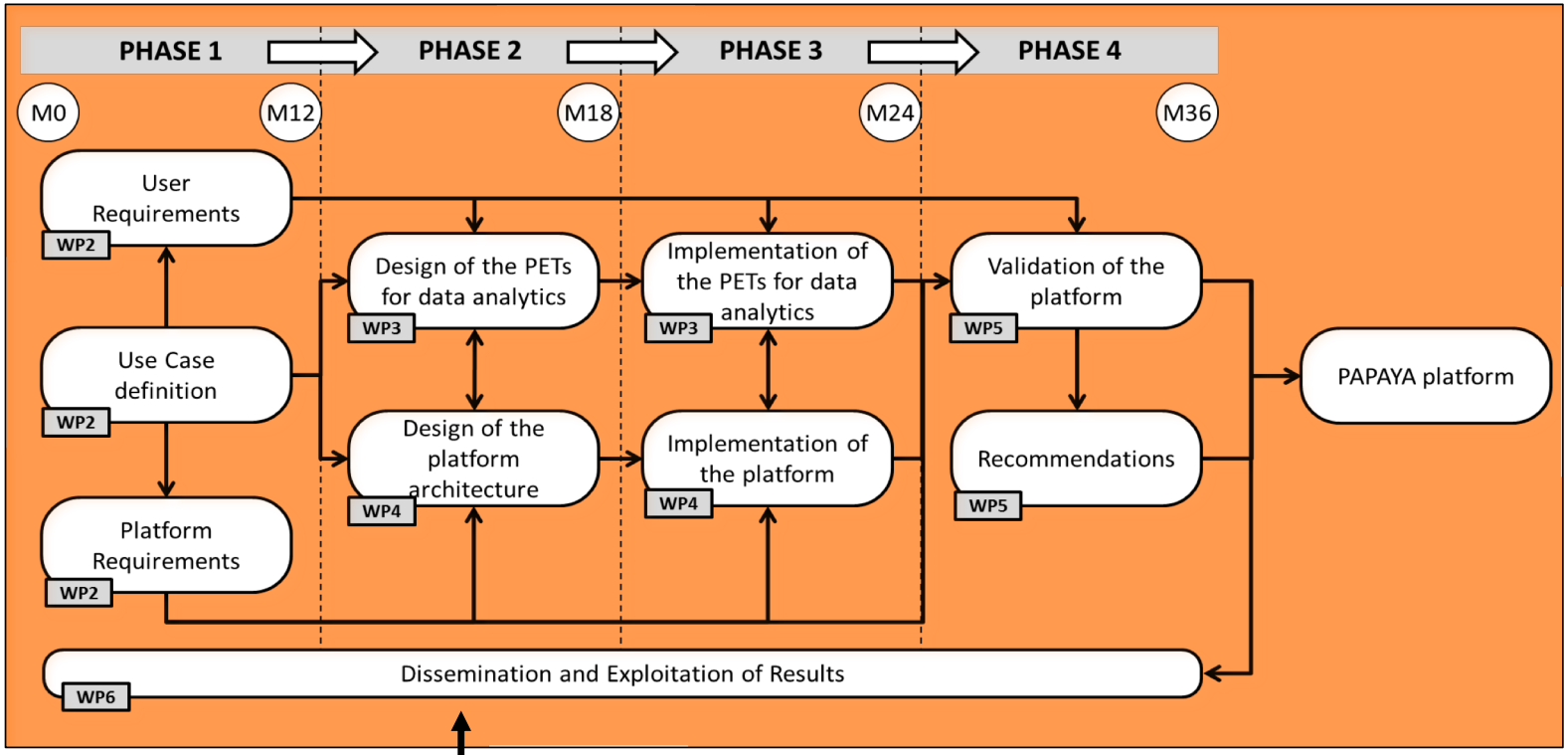


- Objectives
  - Privacy by design
    - PP analytics: processing over protected data
  - Different settings
    - Single vs multiple Dos
    - Third party queriers
  - Integrated platform
    - Common framework
  - User control
    - Transparency, usabilitiy & auditability

PAPAYA

PIAtform for PrivAcY preserving data Analytics

Project Roadmap



PAPAYA Use Cases

[D2.1]

- Healthcare umbrella
  - Arrhythmia detection
  - Stress detection

MEDIACLINICS
Wearable Health Applications

- Mobility and phone usage umbrella
  - Mobility analytics
  - Mobile usage analytics
  - Threat detection

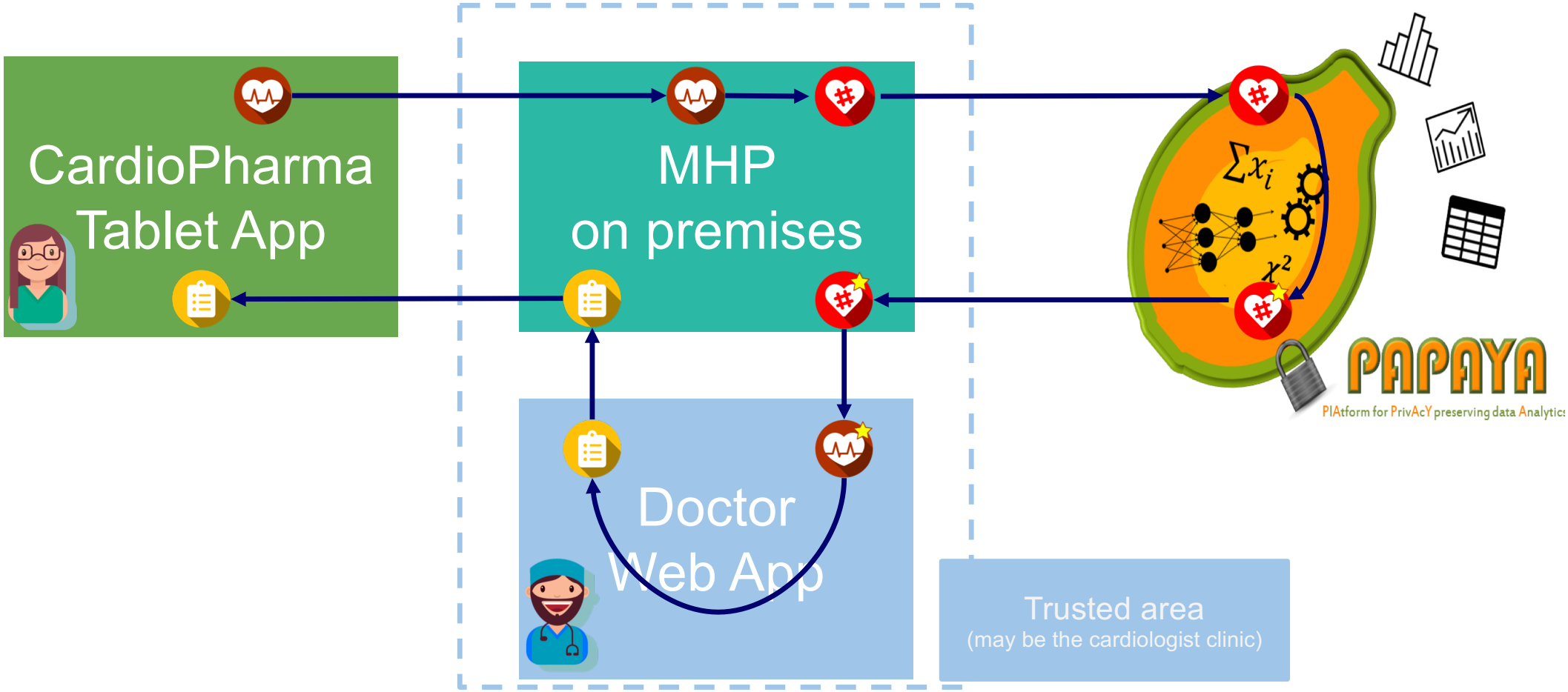
orange™

PAPAYA Analytics

[D3.1]

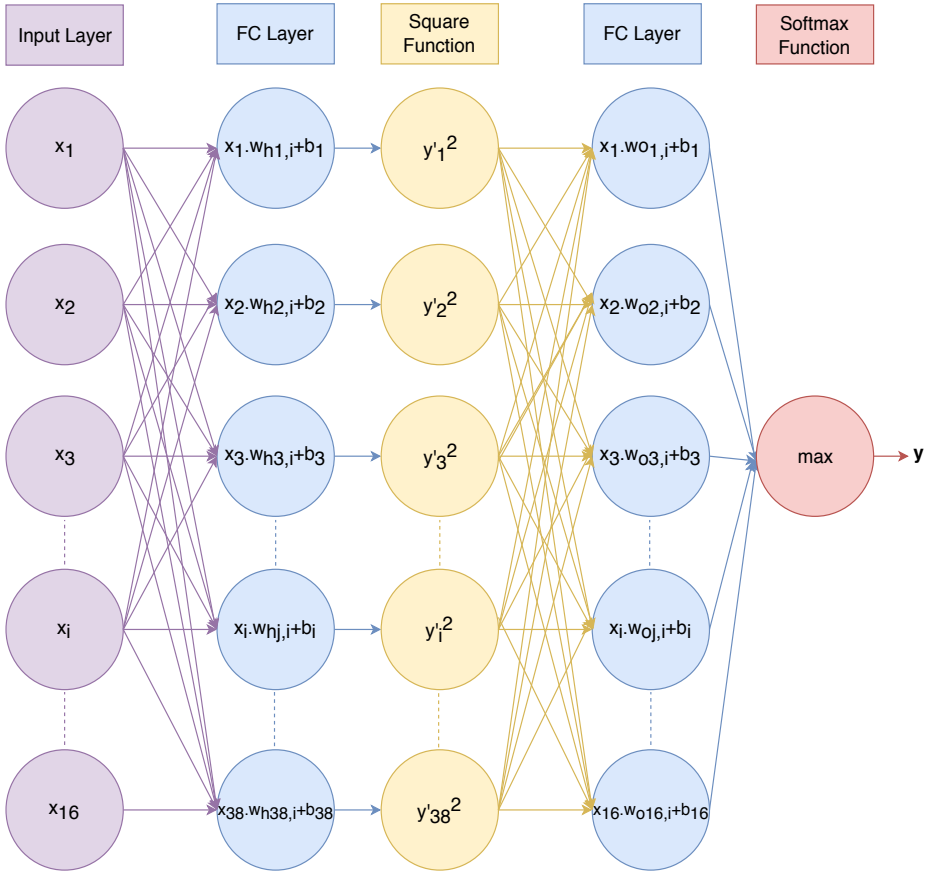
- Neural Network classification
  - Arrhythmia detection, Threat detection
- Collaborative Neural Network training
  - Stress detection
- Trajectory clustering
  - Mobility analysis
- Counting (& and set operations)
  - Mobile phone usage

Arrhythmia detection with Neural Networks



Privacy vs. Neural Networks

- Advanced Privacy technologies
  - FHE, MPC
- Challenges
  - **Additional overhead**: Computation, memory and bandwidth
  - **Complex operations** (sigmoid, tanh, etc.)
  - **Real numbers** (vs. integers with PETs)
- Goal
  - Reduce NN complexity
  - Use low degree polynomials
  - Use integers
  - Keep good level of accuracy
- Solution for PAPAYA
  - ⇒Generate a dedicated NN model from scratch



PP arrhythmia classification based on 2PC

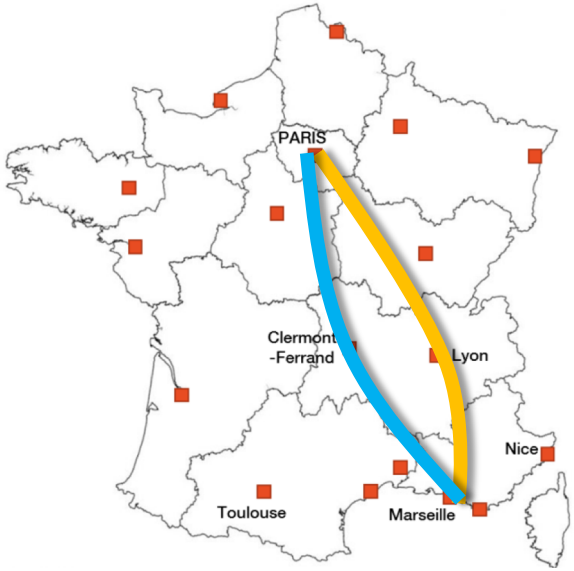
[IFIP Summer School on Privacy'19]

- NN architecture optimization
  - PCA to reduce input size
  - Minimum number of hidden layers with good accuracy
- Approximate non linear operations
  - Square (x^2) instead of sigmoid ($\mathrm{f(t)} = \frac{1}{1+e^{-t}}$)
  - Simple max instead of softmax $\left(\mathrm{f}(i) = \frac{e^i}{\sum_{i=1}^{n} e^i} \; where \; i = 1, \ldots, n\right)$
- Approximate real numbers
  - Truncation: $\times 10^r$
- Performance
  - 62ms to classify one heartbeat, 400 ms to classify 10

Privacy vs. Clustering

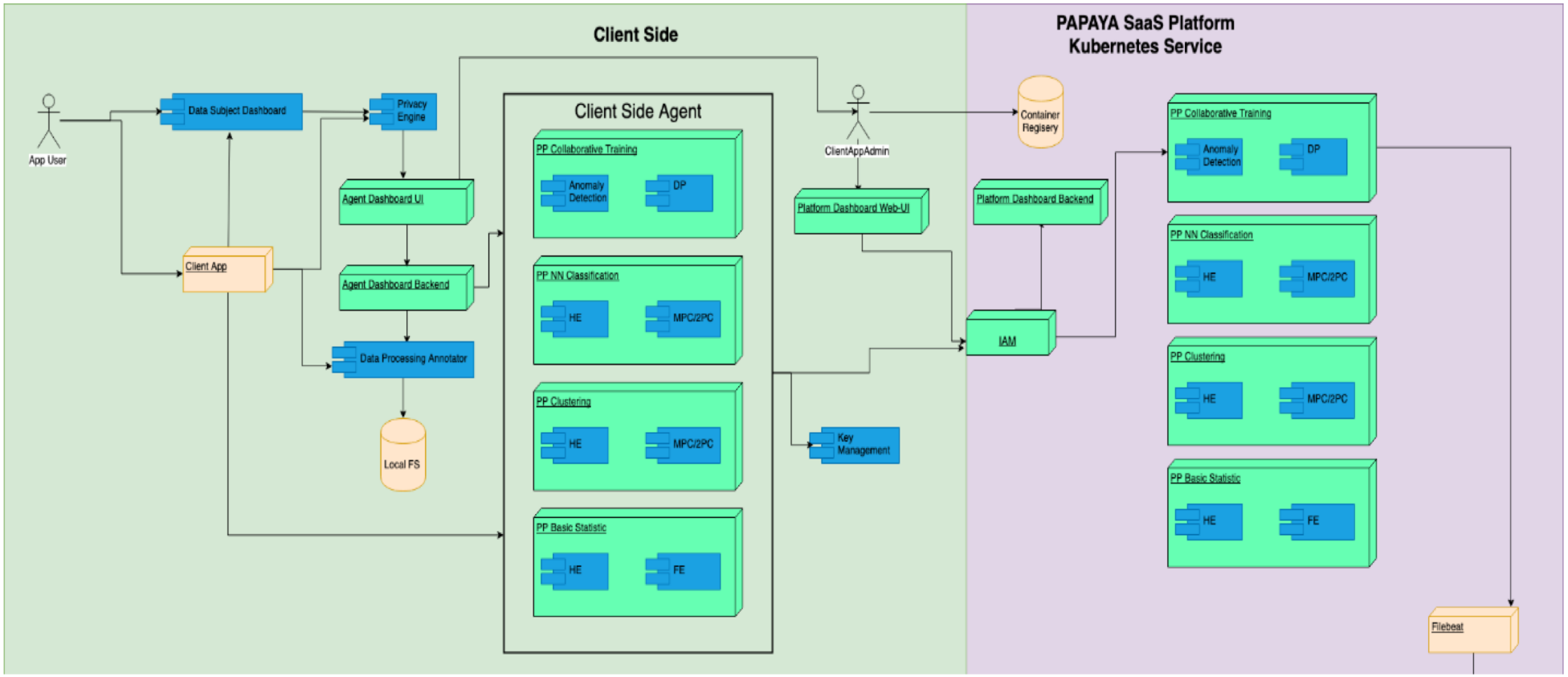
- Objective
  - Encrypted trajectory clustering
- Challenges
  - Processing of Personal data (location, ID, etc.)
  - Sophisticated distance computation
  - Comparisons
- PAPAYA solution

⇒ Approximate distance + use 2PC



PAPAYA Architecture (v1)

[D4.1 available soon]



Thank you

The work described in this presentation has been conducted within the project PAPAYA. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 786767. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

Co-funded by the Horizon 2020 Framework Programme of the European Union

PAPAYA Dashboards

- Platform Dashboard (Web application)
  - Service catalog
  - Service Add/Delete/Update
  - Application Create/Delete/Deploy
  - Application monitoring for service owners
- Agent Dashboard
  - Agent Configuration
  - Data processing logs
- Data Subject Toolbox
  - Explaining PP Analytics
  - Data Disclosure Visualization Tool
  - Annotated Log view tool
  - Privacy Engine