Privacy preserving data analytics
Coimbra, 11 July, 2019
Melek Önen
# PAPAYA Project

<table>
<thead>
<tr>
<th>Project title</th>
<th>PAPAYA: PlAtform for PrivAcY preserving data Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call</td>
<td>DS-08 Cybersecurity PPP: Privacy, Data Protection, Digital Identities</td>
</tr>
<tr>
<td>Grant Agreement</td>
<td>GA no: 786767</td>
</tr>
<tr>
<td>Project Officer</td>
<td>Mr. Nikolaos Panagiotarakis (H2020, REA)</td>
</tr>
<tr>
<td>Duration</td>
<td>36 months (1 May 2018 – 30 Apr 2021)</td>
</tr>
<tr>
<td>Consortium</td>
<td>![EURECOM Logo] ![Atos Logo] ![IBM Logo] ![MediClinics Logo] ![Orange Logo]</td>
</tr>
<tr>
<td>Project Coordinator</td>
<td>EURECOM</td>
</tr>
</tbody>
</table>
Data Analytics as a Service
# 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

![2019 Data Breaches in Europe](https://www.bankinfosecurity.com/gdpr-data-breach-reports-to-eu-exceed-59000-a-12006)

Workshop on Privacy, Data Protection and Digital Identity, July 2019
- 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, usability & auditability
Project Roadmap

Currently at M15

Workshop on Privacy, Data Protection and Digital Identity, July 2019
PAPAYA Use Cases

• Healthcare umbrella
  • Arrhythmia detection
  • Stress detection

• Mobility and phone usage umbrella
  • Mobility analytics
  • Mobile usage analytics
  • Threat detection
PAPAYA Analytics

• 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

CardioPharma Tablet App

MHP on premises

Doctor Web App

Trusted area (may be the cardiologist clinic)
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

- 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 ($f(t) = \frac{1}{1+e^{-t}}$)
  - Simple max instead of softmax ($f(i) = \frac{e^i}{\sum_{i=1}^{n} e^i}$ where $i = 1, ..., n$)
- 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)

Workshop on Privacy, Data Protection and Digital Identity, July 2019
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.

PAPAYA GA Meeting, Madrid, 5-6 March 2019
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