Law, Science and Technology
MSCA ITN EJD n. 814177

On the Efficiency of Decentralized File Storage for Personal Information Management Systems

Mirko Zichichi[1,2], Stefano Ferretti[3], and Gabriele D'Angelo[2]

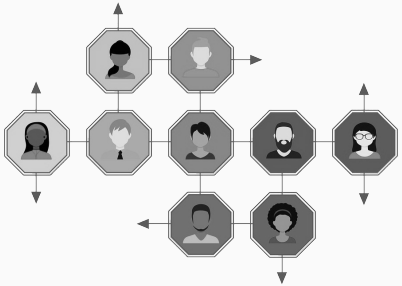
[1]Universidad Politécnica de Madrid
[2]University of Bologna
[3]University of Urbino "Carlo Bo"

Overview

1. Personal Data
2. Distributed Technologies
3. Performance Evaluation
4. Conclusion

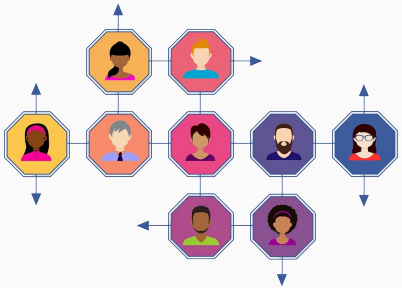
Personal Data

Social Media Personal Data



- **Social media** and Web 2.0 → broke boundaries in **authorship** and **readership**
- [$] **of personal data** is helped by the more **pervasive** nature of today's digital world
- [+] **personalization** ⇒ [+] **privacy threats** for user-generated content
- **Platform-centered** data management ⇒ [-] **transparency** on the use of users' data

Internet of People (IoP)



- **Internet of People (IoP)**:
  - leverages such centralized platforms, when needed
  - places **individuals** at the heart of the **data management** design
- **Smartphones** and personal IoT devices will function as **gateways**

Internet of People (IoP)

- **Internet of People (IoP)**:
  - leverages such centralized platforms, when needed
  - places **individuals** at the heart of the **data management** design
- **Smartphones** and personal IoT devices will function as **gateways**
- **Main issue**:
  publish data while granting compliance with regulations, i.e. **GDPR**

Personal Information Management System (PIMS)

To ensure **sovereignty** of personal data and its **interoperability** we use the:

Personal Information Management System (PIMS) model

a **virtual boundary**, where individuals can **control**

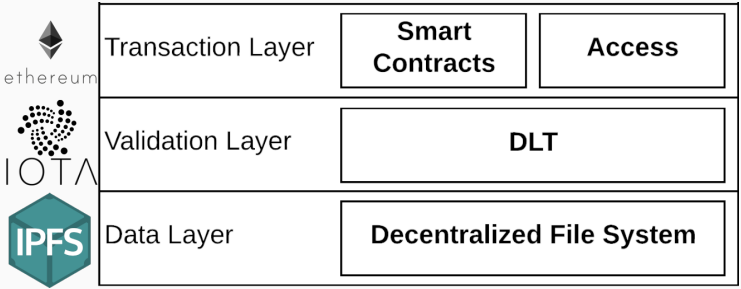
how, *when* and *what* data is shared with external parties

- adheres to **transmission** and **processing** of personal data rules of GDPR
- acts as a strong facilitator for the **consent** of individuals

Distributed Technologies

Decentralized architectures

Decentralized architectures might be the key to foster individuals' data **sovereignty** and fair data **transfer**.



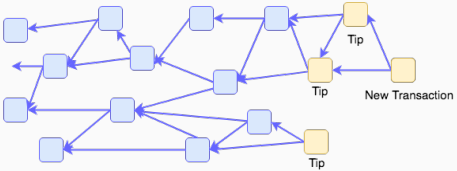
We propose an architecture based on **Distributed Ledger Technologies (DLTs)** and **Decentralized File Storage (DFS)** able to manage personal data storage and access.

Smart Contracts

- "**Trustless trust**" → trust is shifted from a human intermediary to the protocol itself.
- **Ethereum Virtual Machine**
  computes (*quasi-*)Turing-complete programs in a distributed way and permanently stores their input and output on the blockchain.
- **Data Access Control**
  Access to the data can be **purchased** or **allowed by the owner** through dedicated smart contract methods
- **Access Control Lists (ACL)**:
  - represent the rights to access a bundle of data of a consumer
  - an **authorization service** checks the ACL to release encryption keys

IOTA Masked Authentication Messaging Channels

- **IOTA** → network of nodes that holds a distributed ledger where transactions are validated without fees
- **Masked Authenticated Messaging (MAM)** → communication protocol that adds the functionality to emit and access an encrypted data channels over IOTA



IOTA Masked Authentication Messaging Channels

- **IOTA** $\rightarrow$ network of nodes that holds a distributed ledger where transactions are validated without fees
- **Masked Authenticated Messaging (MAM)** $\rightarrow$ communication protocol that adds the functionality to emit and access an encrypted data channels over IOTA
- IOTA (and DLTs in general) offer data **immutability**, **verifiability** and **traceability**
- Personal data (and large sized non-personal data) is referenced in MAM channels through **hash pointers**, in order to exploit those features

IPFS

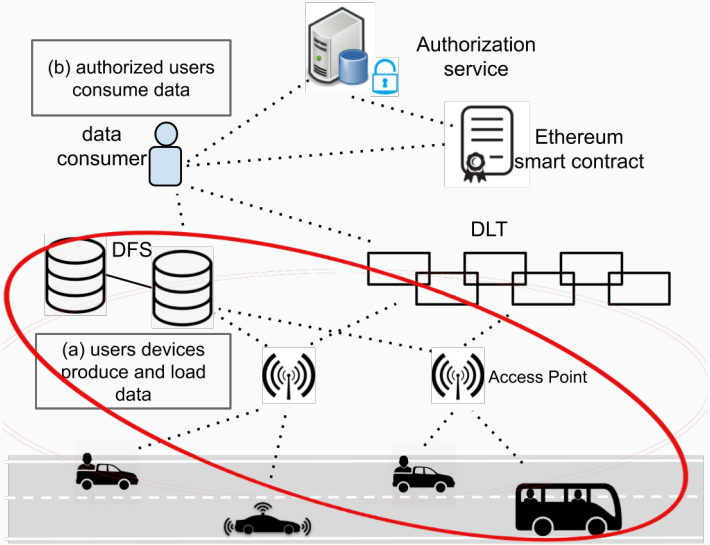
- **InterPlanetary File System (IPFS)**
  - A DFS that creates a resilient file storage and sharing system
  - Useful to store data that is not convenient to put on DLTs
- Once a file is published in the DFS, the **identifier** can be exploited to retrieve it
- Uses **data digest** as identifier ← hash pointer
- **Personal data** → is published as an **IPFS object** → referenced through its hash pointer into a **MAM channel**
- The digest allows verifying the **integrity** of the data

SIA

- IPFS does not offer guarantees on the **persistence** of data
- **SIA**
  integrate a DLT to provide incentives for nodes to maintain data
- **File Contracts**
  agreements between a storage provider and their clients on DLT
- **Skynet**
  nodes that already formed contracts with every available host and providing a service with its own policies

Performance Evaluation

Use case [1/2]


(b) authorized users consume data
Authorization service
data consumer
Ethereum smart contract
DLT
DFS
(a) users devices produce and load data
Access Point

Use case [2/2]

- **Small sized data**: geolocation (100 bytes), encoded as a JSON of this form:

```
{ payload: { latitude: '-22.976509',
             longitude: '-43.19902' },
  timestampISO: '2020-04-05T14:54:11.288Z' }
```

- **Large sized data**: photos (1 MB).

DFS Node Type

1. **IPFS Proprietary**
   - An IPFS node on a dedicated device (dual core CPU, 8GB RAM), connected to other nodes in the main network
   - Receiving requests **only from our test**
2. **IPFS Service**
   - An IPFS service provider (**Infura**)
   - Receiving requests from all over the world (one of the **most used** provider)
3. **Sia Skynet**
   - A Sia node in the **Skynet**, without the needs to create a File Contract
   - Receiving fewer requests than Infura (relatively **new service**)

Sending geolocation to DFS nodes

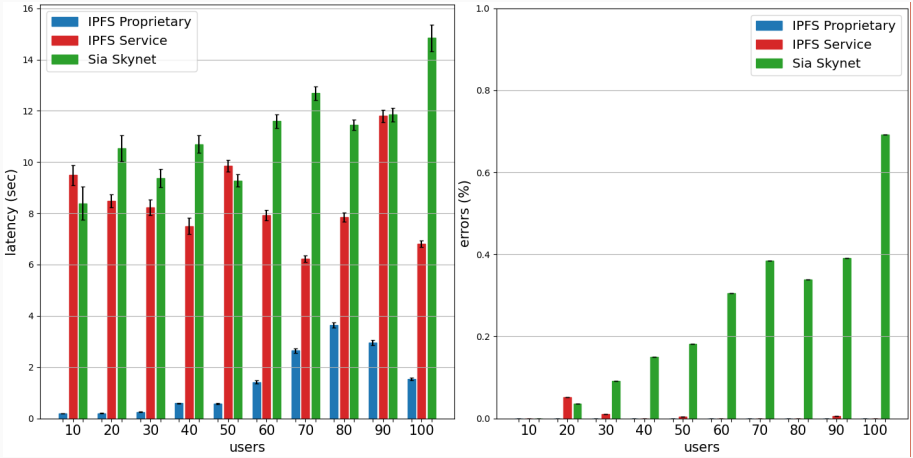


Figure 1: Latencies and errors sending geolocation. Black line → confidence interval (95%)

Sending photos to DFS nodes

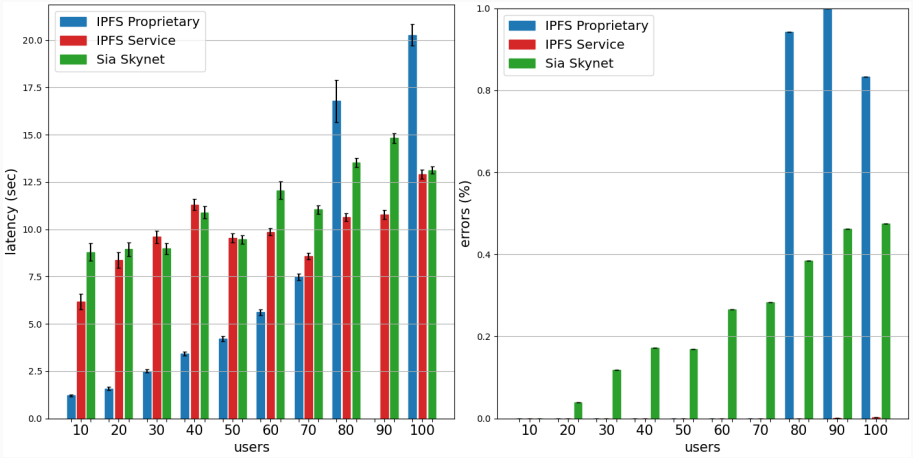


Figure 2: Latencies and errors sending photos (1 MB). Black line → confidence interval (95%)

Conclusion

Conclusion

- **Architecture based on DLTs and DFS** for the development of a decentralized **Personal Information Management System (PIMS)**
- Tested **Infura IPFS**, **Sia Skynet**, and a **proprietary service**
- **Proprietary solution** seems to offer better guarantees in terms of **responsiveness and reliability**
- **Future Work**
  - Further experiments with other scalable DLTs
  - Decentralized authorization service