DARC: Decentralized Anonymous Researcher Credentials for Access to Federated Genomic Data

Mohammed Alghazwi, Fatih Turkmen, Dimka Karastoyanova

University of Groningen

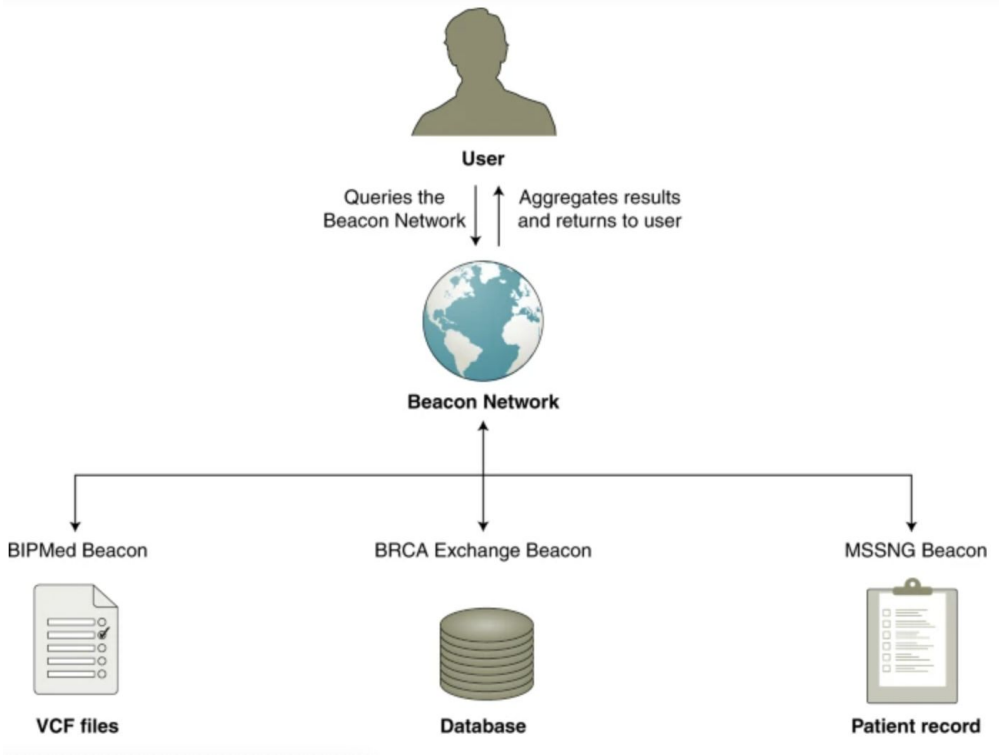
Groningen, Netherlands

m.a.alghazwi@rug.nl

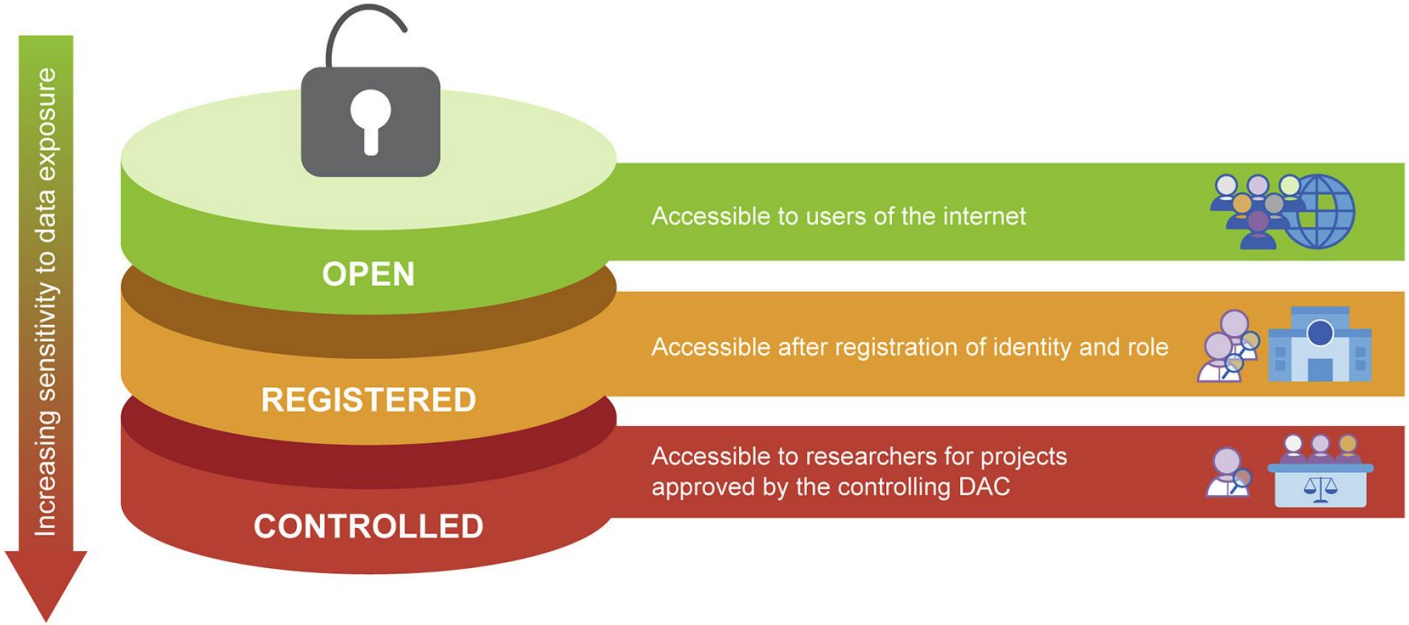
university of groningen

Federated discovery and sharing of genomic data

Genomic Beacon: a simple genomics variant discovery tool by aggregating worldwide genomics dataset under one umbrella.



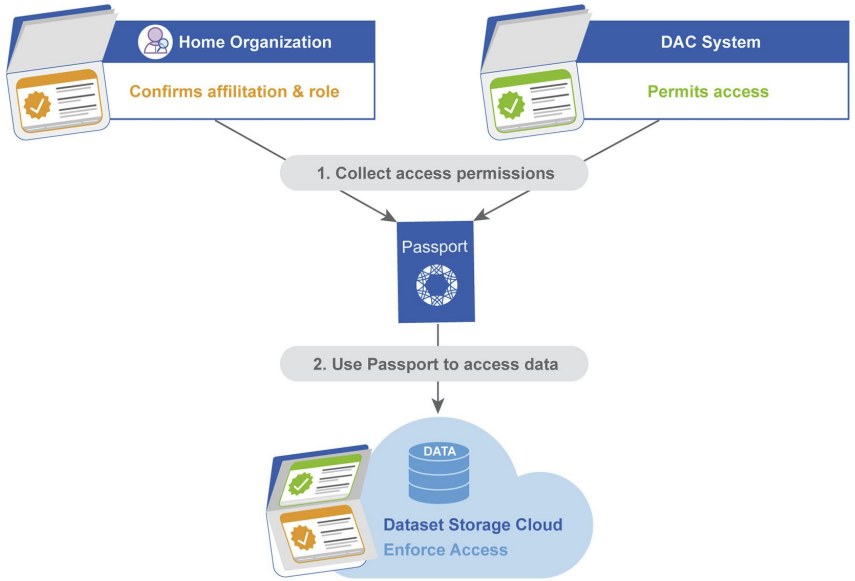
Three-tiered data-access model



Existing Solution

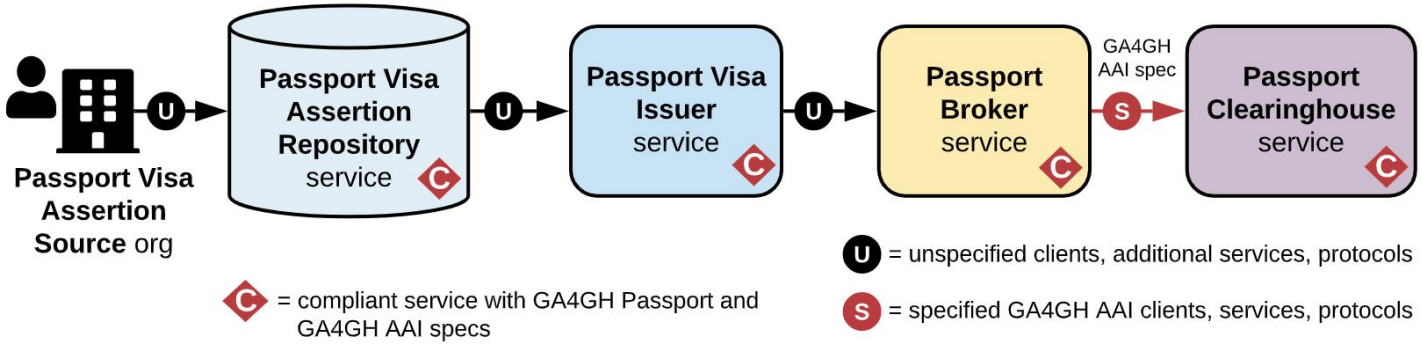
Global Alliance for Genomics and Health (GA4GH) Standards [1]:

- GA4GH passports
- GA4GH visa



[1] Voisin C, Linden M, Dyke SO, Bowers SR, Alper P, Barkley MP, Bernick D, Chao J, Courtot M, Jeanson F, Konopko MA. GA4GH Passport standard for digital identity and access permissions. Cell Genomics. 2021 Nov 10;1(2):100030.

GA4GH Passport and visa standard



Challenges

- Lack of **trust framework** that helps the data repositories (Passport Clearinghouse) to decide which Passport Brokers or Visa Issuers can be trusted
- **Reliance on third-parties** (visa issuer/broker services).
- **Privacy** of the query and **tracking** researcher activities across databases.

Objectives

- Build a **trust framework** between the parties involved in the federated genomic data sharing use case.

- **Aggregate** researcher identity data (issued by different issuers/organizations), in a privacy-preserving way

- **Selectively reveal** identity data to meet the Three-tiered data-access model

Desired Properties

- Only researchers holding a **valid** credential can get access to the data.
- Credentials remains **private** unless selectively revealed by the holder.
- Revealed claims should only reveal the **type of claim** (the group it belongs to), not the specific owner (account) of that claim.

Assumptions:

- CIs are trusted to record valid credentials
- Data repositories follow the protocol in checking the registry for the appropriate credentials prior to providing access.

Methods

- Merkle Trees to represent set membership
- Zero-Knowledge Proofs (ZKP) to prove membership
- Blockchain and Smart contracts for distributed trust

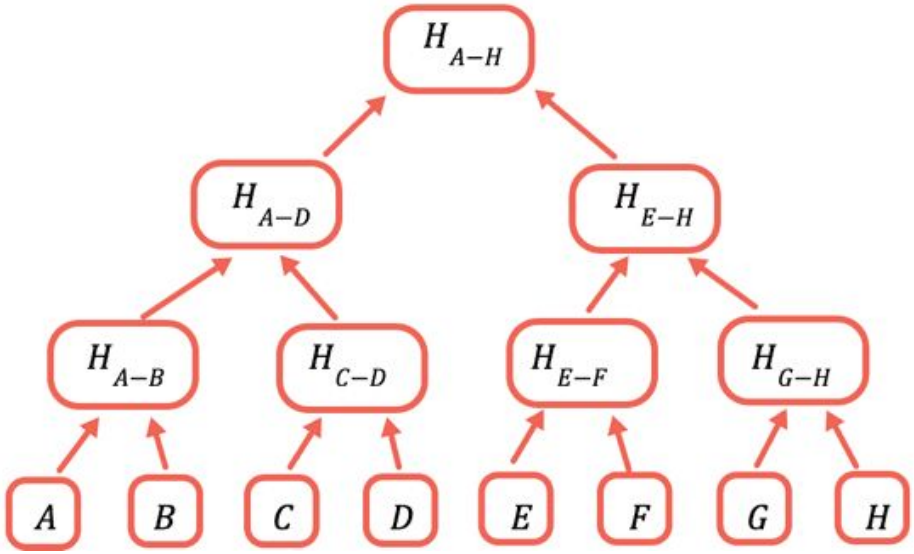
Merkle Tree

- ***Add(k,v)*** → ***MT'***, adds the Poseidon hash of key-value pair H(k,v) to the tree, and outputs the modified tree MT'.

- ***getRoot()*** → ***R***, returns the current root of the tree R.

- ***Prove(k,v)*** → ***α***, given key k and value v, generate the path (proof) α used to prove that H(k,v) ∈ MT.

- ***Verify(k,v,R,α)*** → ***{0,1}***, outputs 1 if the key-value pair is in the Merkle tree, and 0 otherwise.

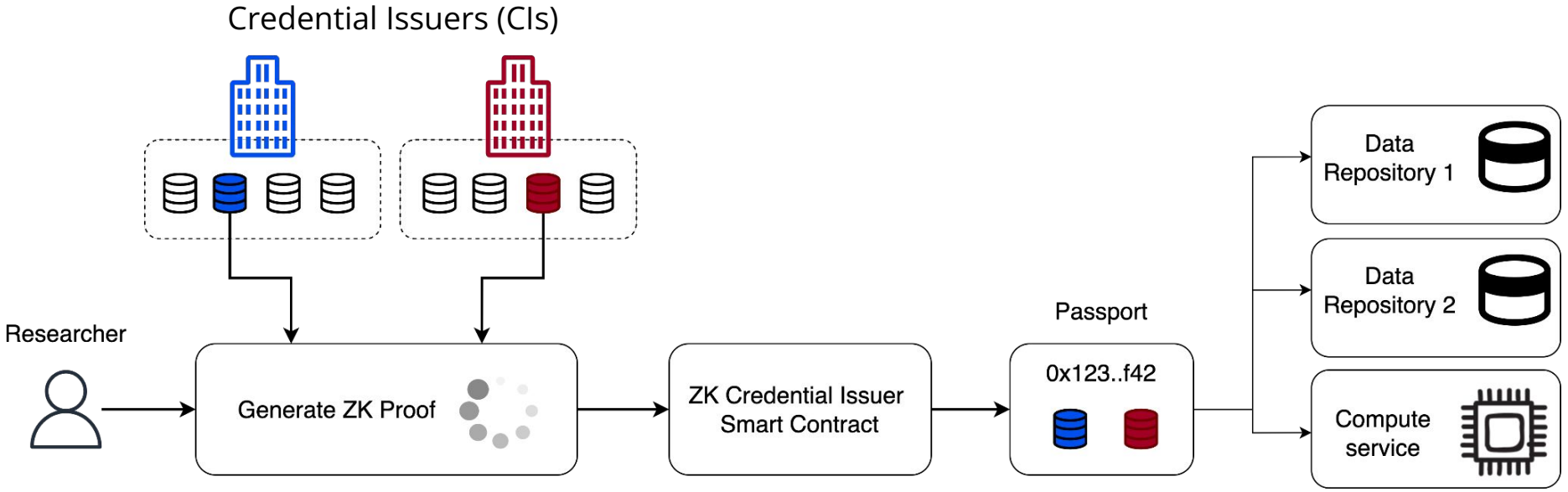


Zero-Knowledge Proofs (ZKP)

- $Setup(1^{\lambda}, \emptyset) \rightarrow crs$
- $Prove(crs, x, \omega) \rightarrow \pi$
- $verify(crs, x, \pi) \rightarrow \{0,1\}$

λ	security parameter
∅	defined circuit
crs	common reference string
x	circuit input
ω	witness (private input)
π	proof

System Model



Group Membership

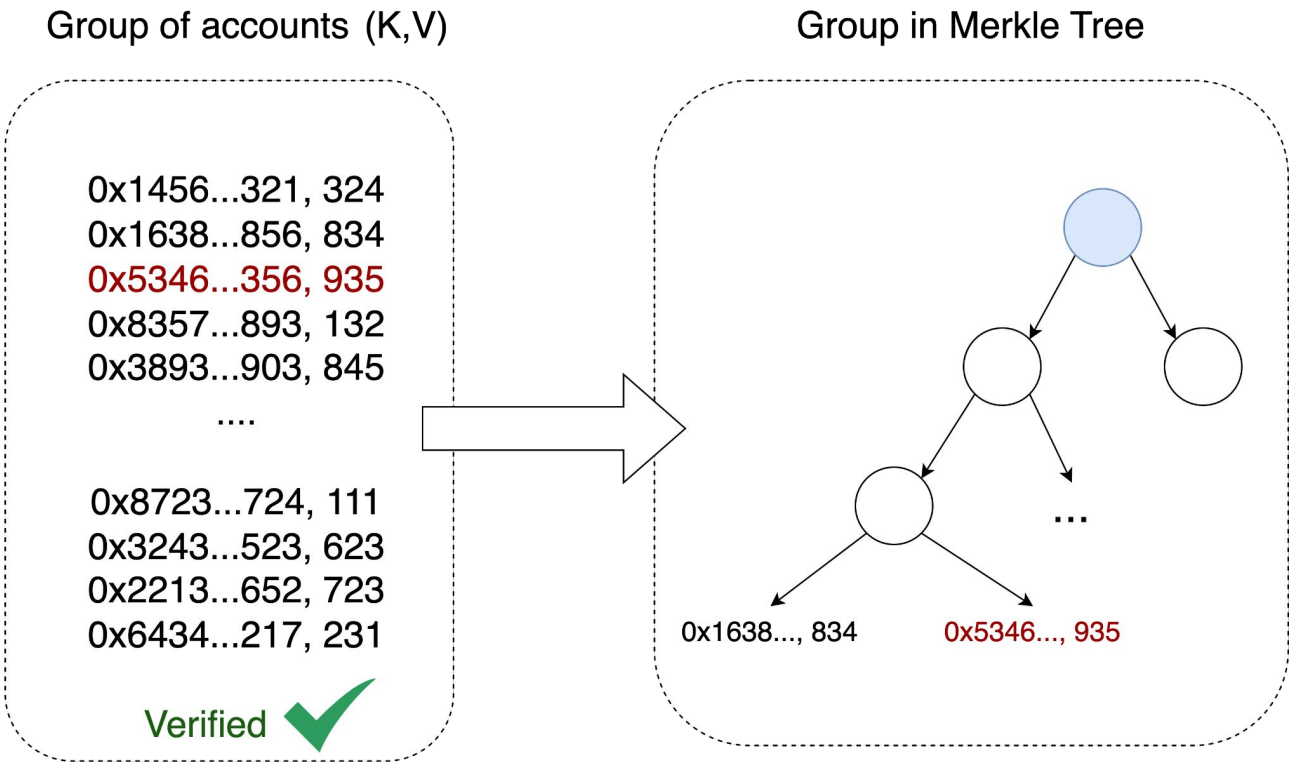
Group of accounts (K,V)

0x1456...321, 324
0x1638...856, 834
0x5346...356, 935
0x8357...893, 132
0x3893...903, 845
....

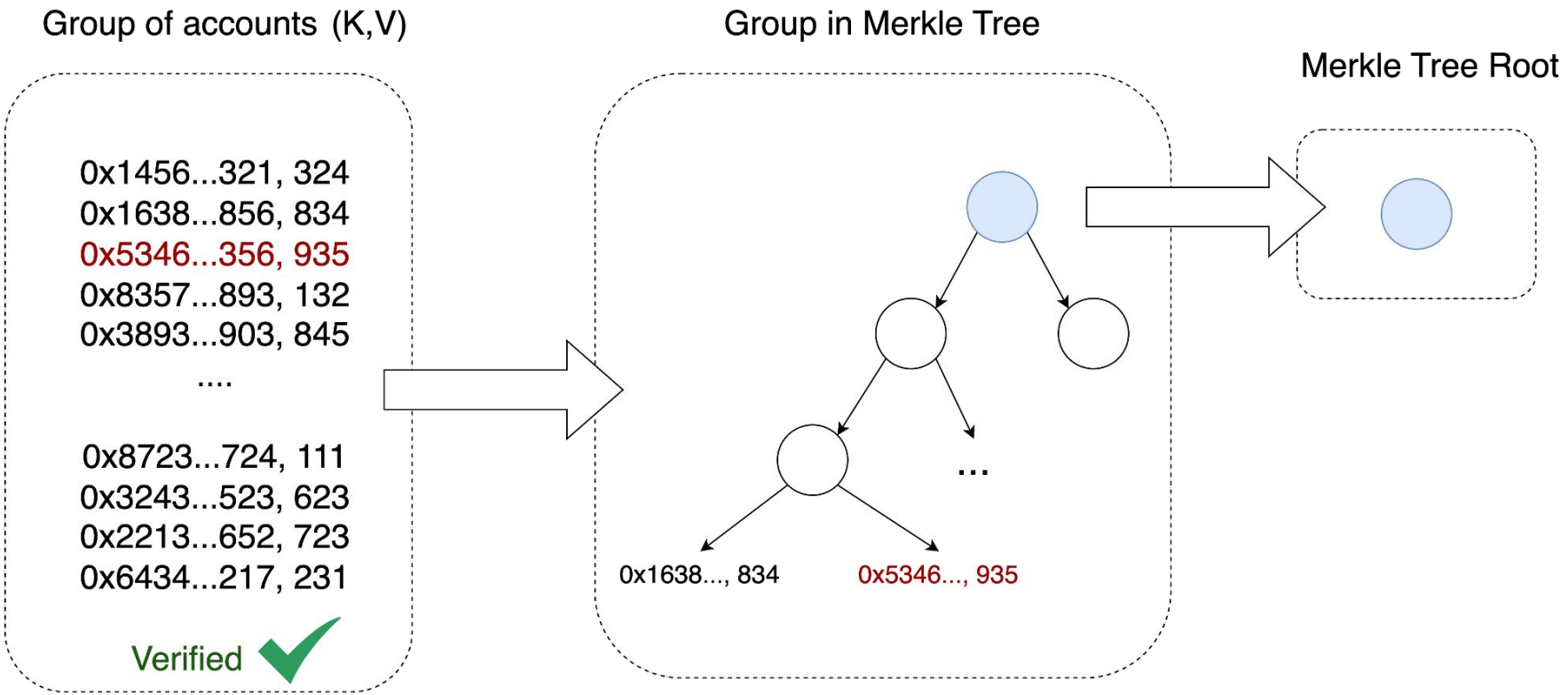
0x8723...724, 111
0x3243...523, 623
0x2213...652, 723
0x6434...217, 231

Verified 

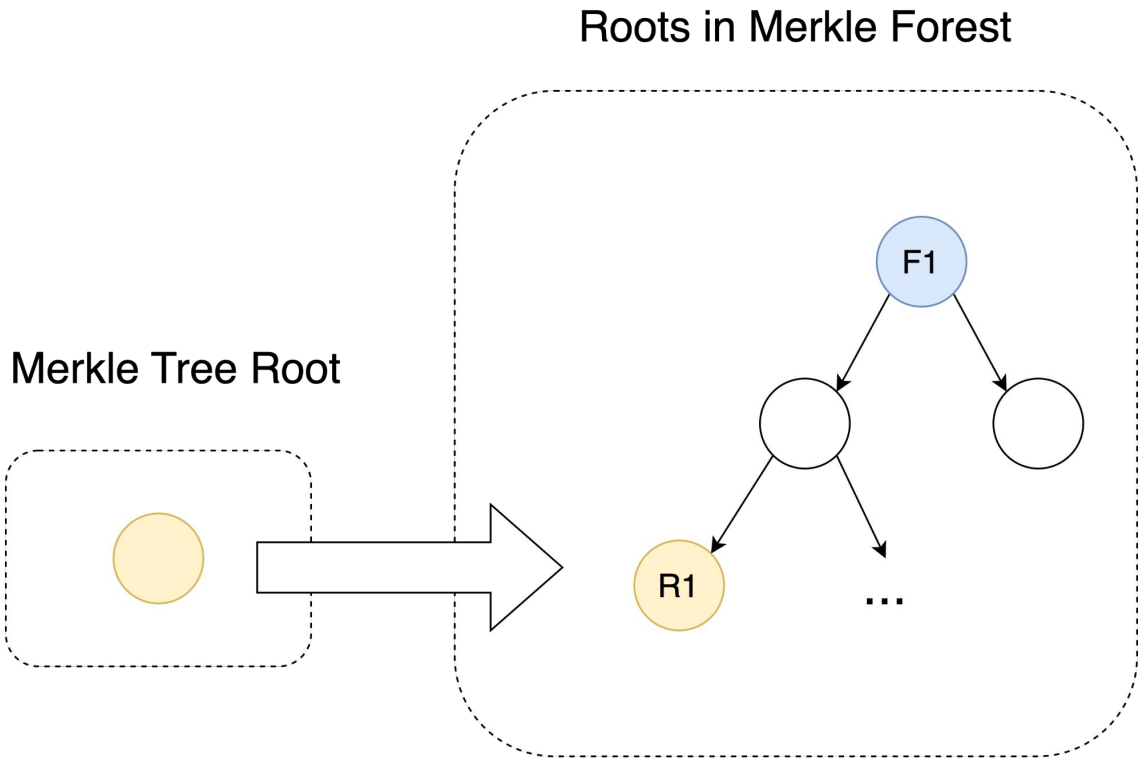
Group Membership



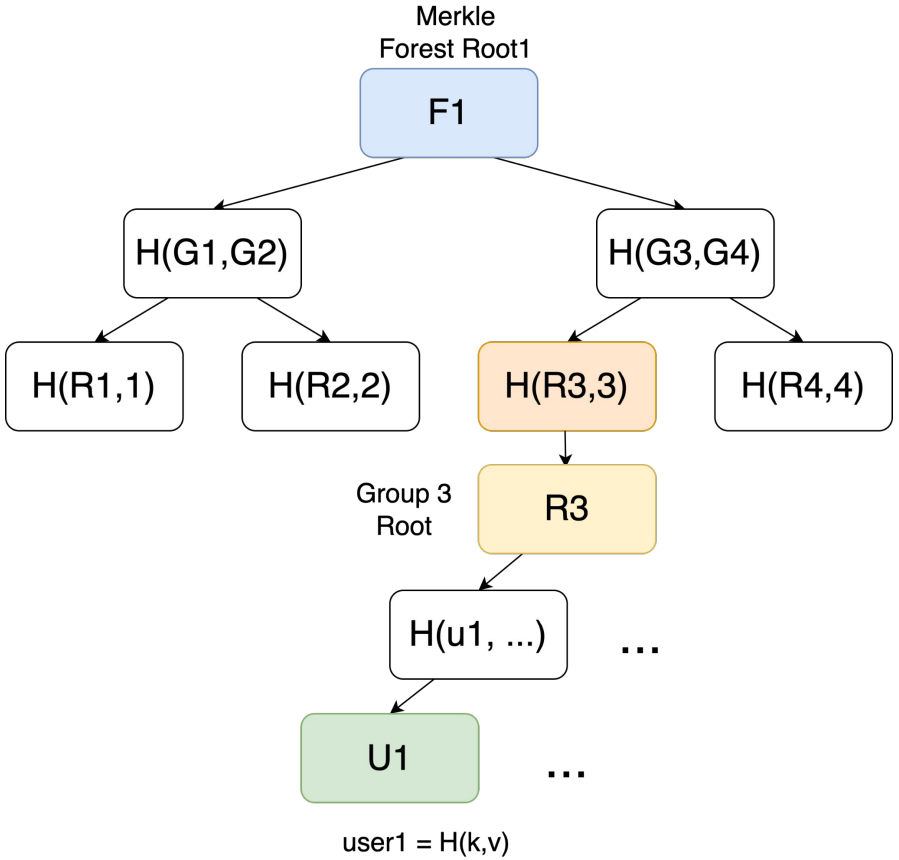
Group Membership



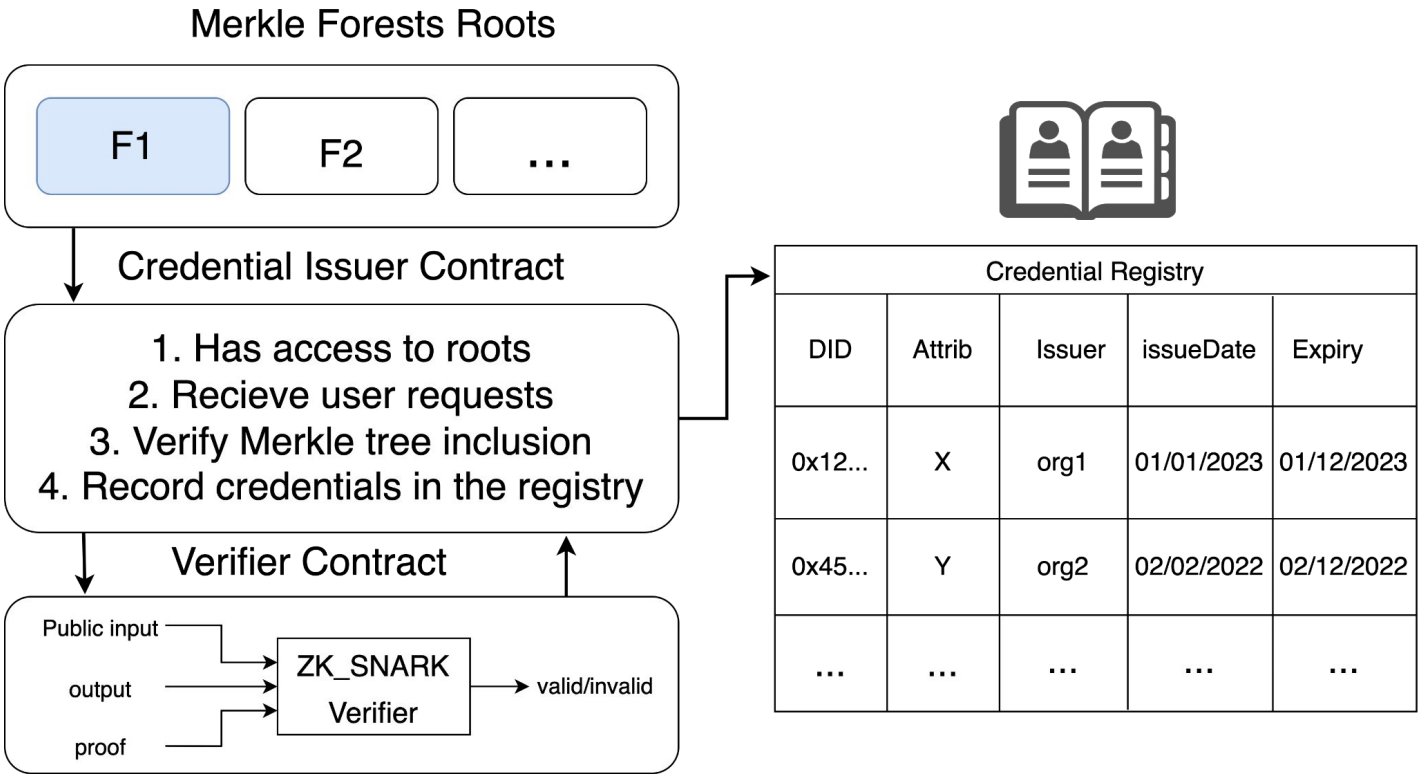
Group Membership



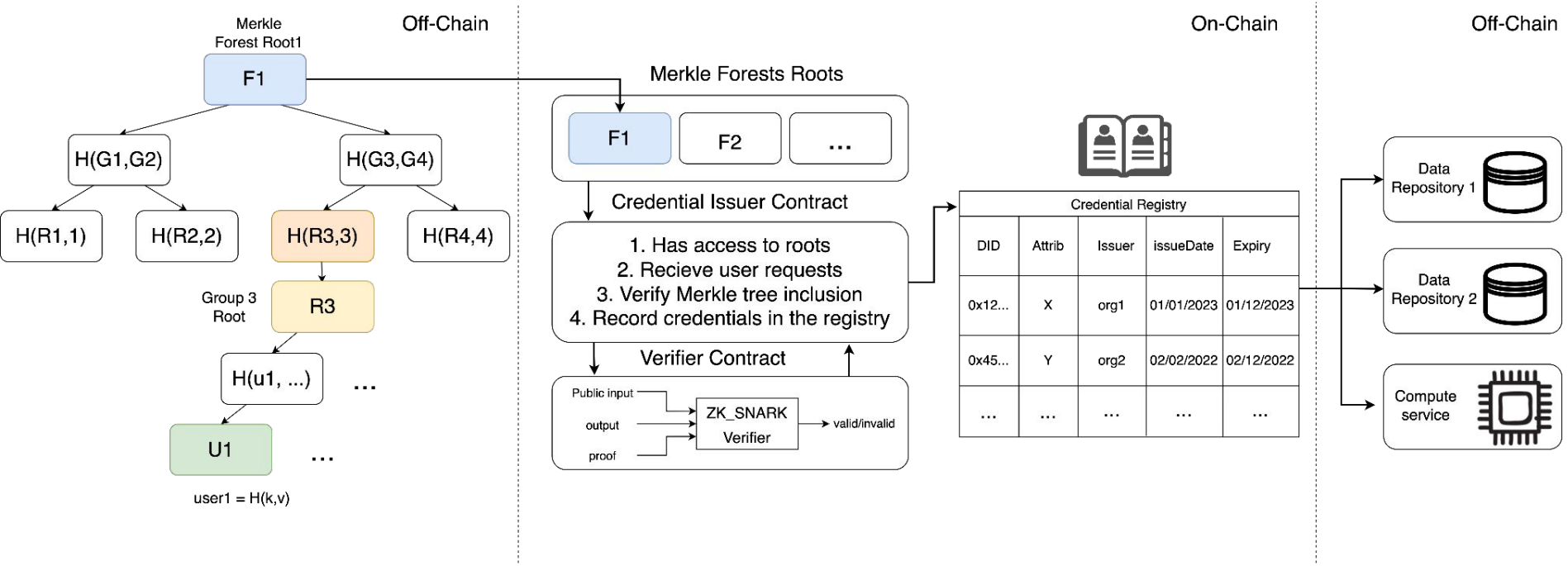
Group Membership



Credential Generation



DARC Protocol Overview



Preliminary Evaluation

Experimental Setup

- “Circom” to define and compile the circuits
- zk-SNARKS scheme: Groth16
- SnarkJS library for executing the compiled circuits
- Poseidon hash function

Metrics

- On-chain cost
  - Gas costs of verification
- Off-chain cost
  - CPU run time: 1.setup, 2.proof generation, 3. verification

Preliminary Evaluation

off-chain Costs

Constraints	10200
Compile time (s)	4.11
Trusted Setup time (s)	27.9
Proving key size (MB)	6.1
Verifier contract size (KB)	12
Proof generation time (s)	2.4

On-chain Costs

Function	Gas Cost
Deploy CI contracts	1,845k
Deploy verifier contracts	1,364k
Store MF root	51k
Verify credential	212k
Store credential	33k

Preliminary Evaluation

Protocol	System Properties		Functionalities			Privacy Guarantees	Additional Properties	
	Distribution of Trust	Modular architecture	Support 3-tier access model	Aggregate claims	Selective disclosure	Identity privacy/ anonymity	User Control	Transparency
GA4GH Passport	×	×	✓	✓	✓	×	×	×
DARC	✓	✓	✓	✓	✓	✓	✓	✓

Future Work

- Evaluate use of side chains or layer 2 blockchains
- Integrate and test DARC with the beacon API
- Credential revocation and Sybil-resistance

Thank You!

Any Questions?

m.a.alghazwi@rug.nl

university of groningen