**Data Minimization**

**Session 1: 4 Oct 2017**

**Daniel Hardman, Evernym/Sovrin**

**Brent Zundel, Evernym/Sovrin**

**Mike Lodder, Evernym/Sovrin**

**Zachary Larson, Independent**

**Irene Hernandez, Independent**

**Lionel Wolberger, Independent**

# Check

## Introduction

Data minimization and selective disclosure (D&S, DataMin, SelDis) are very cool applications of crypto to do magic tricks, such as proving a person is over 25 years old without revealing their birthday or even which credential vouched for the person.

These capabilities are needed for creating, storing, presenting, and verifying user-controlled credentials among other things: DataMin is one of three mitigations against privacy threats in RFC6973, it is featured in article 5, 25 of the GDPR, the USA Privacy Act of 1974 and often appears in FIPPs. This group's goal is to standardize D&S techniques for the Verifiable Claims work (to be used in Blockchain systems supporting self-sovereign identity), an official work item of the W3C Credentials Community Group. Some topics we plan to address include Merkle trees for redaction, Progressive disclosure, CL Signature schemes (Camenisch-Lysyanskaya), ZK (zero knowledge) protocols such as Fiat-Shamir, ZK Snarks and Starks; also commercial providers such as Qredo.

Current known participants in this work item are:

Lionel Wolberger

Jan Camenisch

Maria Dubovitskaya

Christopher Allen

Kim Hamilton Duffy

The overall agenda is to bring a range of intellectual horsepower from cryptography interest/expertise to simply wrapping one's head around what we can (practically) accomplish in this space. We may pop up several levels to think through user stories to clarify the needs.

Logistics

Go wide before we go deep

Go deep on language

Go deep on crypto & tech

Get code

Organize, prioritize

Report

This inventory is a step towards supporting the drafting and incubating of related Internet specifications, as well as further standardization and prototyping and testing reference implementations.

### References

Data Minimization and Selective Disclosure Repo: https://github.com/w3c-ccg/data-minimization

2010 Pfitzmann, Hansen. A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management" Link: https://www.researchgate.net/publication/234720523\_A\_terminology\_for\_talking\_about\_privacy\_by\_data\_minimization\_Anonymity\_Unlinkability\_Undetectability\_Unobservability\_Pseudonymity\_and\_Identity\_Management

RFC6973 Cooper, Tschofenig, Aboba, Peterson, Morris, Hansen, Smith, Janet 2013. Link: https://tools.ietf.org/html/rfc6973. The draft can also be helpful, "This document focuses on introducing terms used to describe privacy properties that support data minimization." 2012 Hansen, Tschofenig, Smith, Cooper. Privacy Terminology and Concepts. Network Working Group Internet-Draft Expires: September 13, 2012. Link: https://tools.ietf.org/html/draft-iab-privacy-terminology-01

Redaction Signature Suite 2016, Draft Community Group Report 26 June 2017. Longley, Sporny. Link: https://w3c-dvcg.github.io/lds-redaction2016/ "This specification describes the Redaction Signature Suite created in 2016 for the Linked Data Signatures specification. It enables a sender to redact information in a message without invalidating the digital signature."

## AGENDA IN MORE DETAILS

Logistics: Who are we, who should we work with, what is our cadence, how do we meet, when is this due.

Go wide before we go deep. What is this? Discovery of other groups working on these issues; literature review. Find standards, FIPPS, trends. Bring a range of intellectual horsepower from cryptography interest/expertise to simply wrapping one's head around what we can (practically) accomplish in this space. We may pop up several levels to see the broader picture.

Go deep on language. Clarify our nouns and verbs: Nouns, Inventory of relevant terms (glossary), the things/algorithms/protocols etc. Verbs; Inventory of relevant use cases, journies, what are we trying to achieve, what the relevant nouns \*do\* for us. How does someone represent what they need, and how can we minimize this or be more selective? Can we handle partial claims e.g. someone asks for proof of zip code?

Go deep on technical issues. What crypto is available? Innovations in the blockchain space? How do we canonicalize (JSONLD)?

Get code: What procedures exist? Is there a low hanging fruit (e.g. Merkle Tree redaction?) Discover or generate examples, sample code, prototypes.

Organize, Prioritize, Bring order: - List the gnarly parts: key management and revocation. Rank solutions for relevance to CCG

Report: Submit a report back to the CCG

# Table of Contents Reading

BACKGROUND

Number Theory Pointers

Factoring: Integer Factorization

Elliptic Curve Problem

Quadratic Residues

Discrete Logarithms

Finite FIeld Arithmetic

MESSAGE PREPARATION

Serialization

File formats, encoding issues (Base64, Base58)

Challenge/Response

Symmetric Cryptography

Limited role in this space

Asymmetric Cryptography

Key Lengths

One-Way Hash Functions

Lengths

SHA3, SHA256

Signatures

Number Generation Issues

Random Generation

Large Prime

Key Management

DID to solve these issues

Secret Splitting

Escrow

Key Recovery

Key Revocation

Multi-Party Computation

Homomorphic Methods

Accumulators (One-Way, Revocation Focus)

Quantum Resistant Crypto

Blinding

Zero-Knowledge Proofs

Generating Keys

Nonlinear Keyspaces

Transferring Keys

Verifying Keys

RSA

ElGamal

Elliptic Curve Cryptosystems

BLind Signatures

Schnorr

Key-Exchange Algorithms

See the DID spec

Zero-Knowledge Proofs of Knowledge

Secure Multiparty Computation

Quantum Cryptography Challenge

Implementations

Secure libraries

Sample code

Efficiency Issues

# Vocabulary

Bearer Token

Pairing space cryptography: Weil

**Data Minimization**

**Progressive Disclosure**

**Selective Disclosure**

Principle of Least Authority

Elliptic Curve Signature Algorithm (ECDSA)

Threshold Signatures

# Cryptographic Flows

## Hash

SHA 256

## Assymetric Key

### RSA

### Diffie Helman

Decisional Diffie Hellman problem

Computational Diffe-Hellman

### ECDSA

## Merkle Trees

## Schnorr Signatures

## BLS Signatures

Type 3 – considered more secure because it is mapping two curves

## Cryptographic Commitments

## RSA

## Paillier cryptosystem

## Damgård–Jurik cryptosystem

## ElGamal

## Witnesses

## ZK

### Fiat Shamir ZK

### ZK Snarks

Speedup of ZK Snarks?

### ZK Starks

## Accumulators

## CL Signatures

## Verifiable Claim Implemented in Sovrin

Jason Law, CTO Evernym

Contributed to Hyperledger Indy

Dmitry Khovratovich, resident cyprotgrapher

Linked data signatures

Jan Camenisch

Anonymous credentials

IDE MIX

Used by IBM for anonymous credentials

This is a form of verifiable claim.

Verifiable Claim Implemented in Sovrin

METHOD

User U:

Generates a [Random Prime]. This is used in traditional Diffie/Hellman RSA type key operations.

User U has a private/public key pair.

Claim A:

Claim A has a pub/priv key pair

Claim provider A signs a set of attributes N..M relating to User U

This Verifiable Claim is signed by User U Private key

A proof of this VC is logged on a public ledger, signed by Private Key of ClaimA

Claim B:

Claim B has a pub/priv key pair

Claim provider B signs a set of attributes N’..M’ relating to User U

This Verifiable Claim is signed by User U Private key

A proof of this VC is logged on a public ledger, signed by Private Key of ClaimB

Ledger:

Has proof of Claim A

Has proof of Claim B

Evaluator E

Accesses the proofs on the ledger

Accesses Claim A with attributes N…M

Independently calculates the verifications and shows that the proofs match the claim.