Self-Sovereign Identity on Public Blockchains and the GDPR

Galia Kondova
University of Applied Sciences and Arts Northwestern Switzerland
Switzerland
galia.kondova@fhnw.ch

Jörn Erbguth
University of Geneva
Switzerland
joern@erbguth.net

ABSTRACT

This paper studies three existing technical solutions for a self-sovereign identity on blockchains and analyzes the arising issues related to the General Data Protection Regulation (GDPR) of the European Union (EU). In particular, the paper provides an overview of the existing Sovrin self-sovereign identity on the Hyperledger Indy public permissioned blockchain as well as uPort and Jolocom on the Ethereum public permissionless blockchain. The paper then concludes with a discussion on the GDPR-compliance of the blockchain-based identity concepts.

CCS CONCEPTS

• **Computer systems organization → Architectures → Distributed architectures → Peer-to-peer architectures**

KEYWORDS

DLT, Blockchain, Self-Sovereign Identity, GDPR, Data Protection

ACM Reference format:

G. Kondova, and J. Erbguth. 2020. Self-Sovereign Identity on Public Blockchains and the GDPR. In *Proceedings of ACM SAC Conference, Brno, Czech Republic, March 30- April 3, 2020 (SAC'20)*, 4 pages. DOI: 10.1145/3341105.3374066

1 INTRODUCTION

The self-sovereign identity (SSI) builds on the notion that the user (the data subject) possesses full control of her data. In particular, the user could store identity data and decide how much data they share. The user could also decide with whom to share their personal data [1].

The implementation of the SSI concept seems to be fully supported by the features of the blockchain technology, in particular, decentralization, peer-to-peer interactions and data integrity [2], [3]. On the other hand, other characteristics of blockchain like the immutability of the data and its storage on public blockchains raise questions as to the compliance of the blockchain data with the General Data Protection Regulation (GDPR) of the European Union. In particular the right of users to rectify and remove data and the identification and obligations of data controllers and processors on the blockchain seem to present "tensions between the GDPR and blockchain" [4].

The paper addresses these issues by starting with an introduction to the SSI concept and its Sovrin, uPort and Jolocom blockchain-based applications. After outlining the major aspects of the GDPR, the paper then provides a discussion on the particular GDPR-compliance challenges associated with each of the technical concepts. The paper concludes with a summary and an outlook.

2 SELF-SOVEREIGN IDENTITY

2.1 The concept

The SSI concept implies that the user (individual or organization) could present their trusted credentials (real-world identity) to a third party without having to engage an intermediary. In the case of a blockchain-based SSI, this process is enabled through the ownership and/or control of certain decentralized identifiers (DIDs) that relate a DID subject to means for trustable interactions with that subject. A DID resolves to a DID document that contains the public key for the DID, any other public credentials, and the network addresses for interaction. The DID document is controlled by the identity owner who possesses the associated private key [5].

There are a number of DID method specifications currently under development on different blockchains including the Sovrin DID Method, ETHR DID Method, Alastria DID Method, Ocean Protocol DID Method, Jolocom DID Method, and others [5]. The paper focuses on the Hyperledger Indy-based Sovrin concept and the Ethereum-based uPort and Jolocom concepts. It is worth noting that the presented concepts are work in progress and the W3C Method Registry should be consulted for any modifications announced after the release of this paper [6].

2.2 Sovrin

The Sovrin SSI is operational on the Hyperledger Indy public permissioned blockchain.

As could be seen from Fig. 1 agents could operate on edge devices such as mobile phones, tablets, laptops, etc. (edge layer). Any private agent with a DID could issue and sign verifiable claims because every DID has an associated public-private key pair. The verifiable claims are exchanged over an encrypted private channel (cloud layer). The verifier then could use the DID of the issuer,

SAC'20, March 30 –April 3, 2020, Brno, Czech Republic

ACM ISBN 978-1-4503-6866-7/20/03.
https://doi.org/10.1145/3341105.3374066

which is usually included in the credential itself, and look up the issuer's public key on the blockchain and verify the signature on the claims (DID layer) [7].

As could be seen from Fig. 1 agents could operate on edge devices such as mobile phones, tablets, laptops, etc. (edge layer).

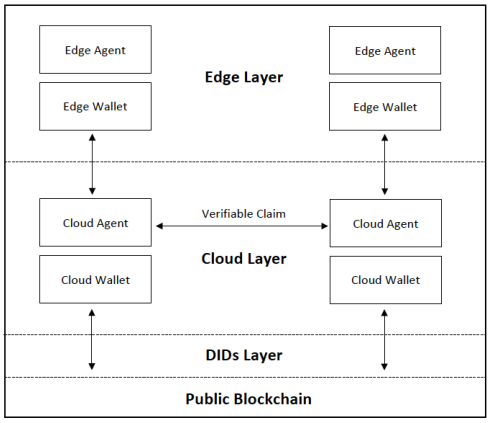

Figure 1: Sovrin Identity Management Architecture

Any private agent with a DID could issue and sign verifiable claims because every DID has an associated public-private key pair. The verifiable claims are exchanged over an encrypted private channel (cloud layer). The verifier then could use the DID of the issuer, which is usually included in the credential itself, and look up the issuer's public key on the blockchain and verify the signature on the claims (DID layer) [7].

The Sovrin ledger (public blockchain layer) stores public DIDs (discoverable entities that belong to organizations or legal entities); the data types and formats used to make up credentials (the schema); "credential definitions" that reference the relevant schema, the issuer who published it and the signature types used; and "revocation registries" (cryptographic numbers maintained by issuers of revocable digital credentials) [8].

2.3 uPort

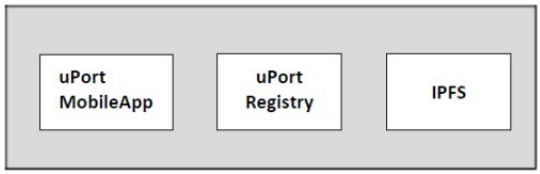

Figure 2: uPort Architecture

The uPort technology is an SSI concept on the Ethereum public permissionless blockchain. According to the uPort Specs, the platform currently consists of the uPort mobile app, Ethereum smart contracts and open protocols for signed messages [9].

The uPort Architecture is presented in Fig. 2. The uPort users manage their identities through the uPort Mobile App (the identity wallet). A request for disclosure is typically signed by a client app and sent to the mobile app. Most request and responses are performed privately off-chain between the different parties. Most off-chain messages consist of signed JWTs (JSON Web Tokens). Any signed message contains a Decentralized ID. Signatures are verified using the uPort Registry [10]. The uPort Registry is a contract which is used to link attributes to identities. Through the uPort Registry, the Interplanetary File System (IPFS) hash for the identity can be looked up. The decentralized identity document (DID Document) stored on IPFS can then be accessed. The DID document contains the public key for the identity. By extracting the public key from the DID Document, the signature verification is completed.

2.4 Jolocom

The Jolocom framework [11] as seen in Fig. 3 by default also stores DIDs on the public permissionless Ethereum blockchain. DID documents (DDO) describe how to use a specifc DID and may contain additional attributes. By default, DDOs are stored on the IPFS. Credentials are under the entire control of the user. Jolocom allows for the generation of child DIDs that can hide that credentials concern the same person.

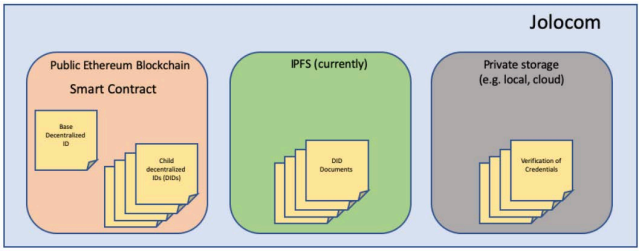

Figure 3: Jolocom storage

Like uPort, Jolocom also comes with a wallet that supports local storage of credentials and key recovery.

3 THE EUROPEAN GENERAL DATA PROTECTION REGULATION (GDPR)

The GDPR has had to be applied in the European Economic Area since May 25, 2018. The GDPR [12] regulates only the processing of personal data – however personal data is subject to a broad definition. The GDPR distinguishes between controllers, processors and data subjects. While the data subjects are attributed rights and protection, the GDPR imposes obligations and sanctions onto controllers and processors. Controllers determine the purposes and means of the processing whereas processors process data on behalf of a controller (Art. 4.7 and 4.8 GDPR). A processing can have multiple controllers. When they jointly determine the purposes and means of the processing, they are considered joint controllers and face special regulations (Art. 26 GDPR).

3.1 How does GDPR apply to SSI?

GDPR applies only to personal data. Personal data is defined as information relating to an identified or identifiable natural person (Art. 4.1 GDPR). In this paper we focus on four types of data in SSI:

DIDs, credentials, revocations of credentials and hashes of one of the first three or other objects. When evaluating whether a person can be identified with certain data, all the means reasonably likely to be used should be taken into account (Recital 26 GDPR). This includes, for example, third parties that have received a credential and kept a copy. To be considered personal data, however, the data needs to convey some (even trivial) information about the data subject – directly or indirectly through its context.

According to this definition, credentials and revocations (if available) are usually personal data, e.g. when the issuing authority and/or the validating entity are able to identify them with the data subject. The situation is more difficult for DIDs and hashes of personal data:

DIDs are not created by some authority but can be created by data subjects. The data subjects prove control of a DID by signing with a private key that is linked to the DID. Although DIDs are related to data subjects, they do not allow the identification unless their usage discloses the identity of the data subject. When DIDs are used only once, this disclosure might be limited to the information that was disclosed and does not link to additional data. However, when used multiple times, the DID works like an identifier and allows to link different credentials or attributes. The same holds true, if some metadata like the time of the creation of the DID discloses z additional information or can be used to correlate other data [14, p. 430].

The systems reviewed also use hashes of credentials or other objects. To understand whether these hashes need to be considered personal data, a case-by-case analysis is required [4, p. 22]. A cryptographic hash function is a non-reversable one-way function. It is not possible to calculate the hashed object from the hash. However, when the hashed object does not contain enough entropy, when the hashed object is partially known and the unknown part does not contain enough entropy, it is possible to brute force the original object by trying every possible value. The hash function can also be reversed, when the hashed object is indexed by its hash in systems like IPFS. The Spanish aepd together with the European EDPS provide advice how hash functions should be applied [13].

However, even when the hashed object has enough entropy, the hash function still works in the direction from the hashed object to the hash value. If the hashed object can be identified with a data subject, this will also enable the identification of the data subject with the hash value. In order to be considered personal data, however, it is not enough to identify a hash value with a data subject. The context of the hash value also needs to convey additional information that is not contained in the object hashed [14, p. 656], [15 p. 430]. This is for example the case, when two hashes are written side by side in order to reference each other.

3.2 Who is controller of DIDs, credentials and revocations of credentials?

A controller is the person or entity responsible for the compliance with the obligations of the GDPR. Art. 4.7 GDPR defines the controller as the natural or legal person that determines the purposes and means of the processing. The control might be based on explicit legal competence, implicit (legal) competence and factual competence [16, p. 9]. There may be several controllers and if several controllers jointly define the purposes and means of processing, controllers might be regarded as joint controllers (Art. 26.1 GDPR). To determine the controller, we distinguish between the data not stored on a blockchain and data stored on a blockchain.

Sovrin, uPort and Jolocom do not store the credentials on a blockchain. The credentials are stored with an agent, a cloud provider or under direct control of the data subject. The controller could be the issuing institution, some agent or the data subject. Usually credentials will only be issued on the demand of the data subject and the data subject can demand the deletion of a credential under her control, whereas the issuing institution can only add a revocation entry. Depending on the SSI system and the use-case the data subject might herself be the controller for adding a credential to an SSI. Yet the roles are different when it comes to revocation of credentials by the issuer which is only supported by Sovrin and uPort. The data subject often has no means to influence revocations issued. Therefore, it is likely that the issuer is considered to be the controller of revocations added to an SSI. When a data subject is herself considered controller of the processing of her personal data, the GDPR does not apply to her, but it may apply to processors that process the data on behalf of the data subject [17, p. 563].

DIDs are stored on a blockchain. On blockchains we have to distinguish between the controller on the blockchain level, the controller on the transaction level and - if applicable - the controller on the smart contract level [15, p. 431]. For public blockchains - as they are used by uPort and Jolocom, the French data protection authority CNIL does not consider node operators and miners as controllers but accepts the possibility, that there is no controller on the blockchain level [18, pp. 2-4]. For a permissioned blockchain, as used by Sovrin, the entity or consortium determining the permissions, might be considered to be the controller. Regarding the transaction level, the person signing a transaction with her private key and sending it to a blockchain might be considered a controller [18, p. 2], [19, p. 564]. When a user sends a privately motivated transaction to a public blockchain, the CNIL holds that the household exemption (Art. 2.2.c GDPR) might apply [18, p. 3]. However, it is unclear, how this can be distinguished from the Lindqvist ruling of the European Court of Justice where the household exemption was excluded for something published on social media [20, para. 47]. On the smart contract level, the CNIL considers a smart contract developer to be a controller (or processor), when the developer has a role in the processing [18, p. 3]. A mere code provider or protocol developer, however, is not considered to be a controller.

As a result, determining the controller of the DIDs and hashes stored on a blockchain requires a case-by-case analysis. It is likely that the data subject will be considered a controller in some cases, while it is also likely that issuing authorities or the entity controlling a permissioned blockchain might be considered to be controller in other cases. The household exemption might apply in some cases. The protocol developer, however, will not be considered to be a controller.

3.3 Justification

Any processing of personal data requires a justification. Possible justifications are given in Art. 6.1 a-f GDPR. Concerning credentials and revocation, there might be consent (Art. 6.1.a) or contract (Art. 6.1.b). A revocation might be required by law (Art. 6.1.c) or justified by a legitimate interest (Art. 6.1.f) to indicate that a credential that has been issued is no longer valid. As a result, justifications depend on the specific use-case and also require a case-by-case analysis.

3.4 Right to erasure / right to be forgotten

A controller has several obligations towards the data subject. Often, the most problematic obligation in the context of blockchain-based systems is the right to erasure / right to be forgotten (Art. 17 GDPR). Off-chain data can be deleted. When the data subject is the controller herself, the right to erasure does not apply. The right to erasure neither applies, when the controller has a justification to continue to store the personal data. Data only referring to issuing institutions usually do not constitute personal data. However, when DIDs, hashes of credentials or hashes of revocations are stored on a Blockchain the above-mentioned case-by-case analysis is required, to determine if some entry (still) needs to be considered personal data. For example, in some cases a hash value which was considered personal data, could stop being considered personal data, when the hashed object has been securely deleted.

3.5 Revocations of credentials

A revocation usually is considered personal data. Therefore, a legal basis is required. In practice, the data subject may not always consent, and a different justification is needed. Dependent on the use-case, there might be a legal obligation, legitimate interest or another legal basis for providing the information about a revocation to people to whom the credential was presented. By linking the revocation to the hash of the credential, the revocation can only be identified with the data subject when the revoked credential is present. The revocation thereby seems to be optimally minimized to be available to only those towards whom there is a legal basis to disclose the revocation.

4 CONCLUSION

Self-Sovereign Identity (SSI) involves personal data. A detailed analysis of the system used and the use-case is required to determine what data components of the SSI constitute personal data, how the GDPR applies and who is considered to be a controller and what justifications exist. When storing some data on an immutable blockchain, it has to be ensured, that either the data stored on a blockchain will not or no longer constitute personal data, that the data subject is considered to be the controller, that the household exemption applies or a permanent justification for continuous storage on the blockchain exists. In many cases, according to Art. 35 GDPR, a data protection impact analysis (DPIA) will be required [18, p. 8].

DIDs, when used for more than one credential, might be considered identifiers that link these credentials. When both credentials are presented, the recipient is able to see that both of them belong to the same person. Although this can be a risk to the data subject and might void some benefits from the use of SSI, it often still protects the data subject better than the use of classic certificates where every certificate is linked directly to the identity of the data subject and thereby also to other certificates that bear the name of the data subject.

Special attention has to be attributed to revocation of credentials. A revocation does not mean the deletion of the credential. It rather adds a revocation entry. This requires a legal basis. Depending on the use-case, this legal basis would often exist (but not always).

SSI can provide a high standard of privacy protection. No central entity has control over the credentials issued. SSI can technically protect the privacy of data subjects and can be compliant with GDPR. However, the requirement of a case by case analysis and the existing legal uncertainty creates a burden to the use of this privacy enhancing technology.

REFERENCES

[1] EU Blockchain Observatory and Forum, “Blockchain and Digital Identity,” 2019. Available at https://www.eublockchainforum.eu/reports (All links accessed December 2019)

[2] F. Zbinden and G. Kondova, “Economic Development in Mexico and the Role of Blockchain,” *Advances in Economics and Business*, vol. 7, no. 1, pp. 55–64, Jan. 2019.

[3] D. He *et al.*, “Virtual Currencies and Beyond: Initial Considerations,” *IMF Staff Discussion Notes*, vol. 16, no. 03, p. 1, 2016.

[4] *EU* Blockchain Observatory and Forum, “Blockchain and the GDPR,” 2019. Available at https://www.eublockchainforum.eu/reports

[5] World Wide Web Consortium (W3C), “Decentralized Identifiers (DIDs) v1.0: Core Data Model and Syntaxes,” 2019. Available at https://w3c.github.io/did-core/#did-document

[6] World Wide Web Consortium (W3C), “DID Method Registry”, 2019. Available at https://w3c-ccg.github.io/did-method-registry/

[7] Sovrin Foundation, “Sovrin: A Protocol and Token for Self-Sovereign Identity and Decentralized Trust,” 2018. Available at https://sovrin.org/library/sovrin-protocol-and-token-white-paper/

[8] A. Tobin, “Sovrin: What Goes on the Ledger?,” Evernym White Paper, 2018. Available at https://www.evernym.com/wp-content/uploads/2017/07/What-Goes-On-The-Ledger.pdf

[9] uPort Specs, 2019. Available at: https://github.com/uport-project/specs

[10] uPort PKI, 2019. Available at: https://github.com/uport-project/specs/blob/develop/pki/index.md

[11] Jolocom, A Decentralized, Open Source Solution for Digital Identity and Access Management, Whitepaper 2.1, December 2019. Available at https://jolocom.io/wp-content/uploads/2019/12/Jolocom-Whitepaper-v2.1-A-Decentralized-Open-Source-Solution-for-Digital-Identity-and-Access-Management.pdf

[12] General Data Protection Regulation (GDPR), 2016. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[13] aepd EDPS, Introduction to the Hash Function as a Personal Data Pseudonymisation Technique, October 2019. Available at https://edps.europa.eu/sites/edp/files/publication/19-10-30_aepd-edps_paper_hash_final_en.pdf

[14] J. Erbguth, "Datenschutzkonforme Verwendung von Hashwerten auf Blockchains," *Multimedia und Recht* 2019, no. 10, pp. 654-660.

[15] J. Erbguth, "Five Ways to GDPR-Compliant Use of Blockchains," *European Data Protection Law Review* 2019, no. 3, pp. 427-433.

[16] Article 29 Working Party, "Opinion 1/2010 on the concepts of 'controller' and 'processor'," 2010, 00264/10/EN WP 169, 9, endorsed by the European Data Protection Board on 25 May2018. Available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169_en.pdf

[17] B. G. Rauber, “Grzeszick/Rauber: Anwendbarkeit der DS-GVO durch Einschaltung Dritter?,” *Zeitschrift für Datenschutz*, no. 12, pp. 560–564, 2018.

[18] CNIL, “Premiers éléments d’analyse de la CNIL - Blockchain,” Sep-2018. Available at https://www.cnil.fr/sites/default/files/atoms/files/la_blockchain.pdf

[19] J. Erbguth and J. Galileo, “Erbguth/Fasching: Wer ist Verantwortlicher einer Bitcoin-Transaktion?,” *Zeitschrift für Datenschutz*, no. 12, pp. 560–565, 201.

[20] European Court of Justice, Case C-101/01 Lindqvist (2003) ECR I-12971