A survey on blockchain-based identity management and decentralized privacy for personal data
Komal Gilani, Emmanuel Bertin, Julien Hatin, Noel Crespi

To cite this version:

HAL Id: hal-02650705
https://hal.archives-ouvertes.fr/hal-02650705
Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A Survey on Blockchain-based Identity Management and Decentralized Privacy for Personal Data

Komal Gilani
Orange Labs, France
komal.gilani@orange.com

Emmanuel Bertin
Orange Labs, France
emmanuel.bertin@orange.com

Julien Hatin
Orange Labs, France
Julien.hatin@orange.com

Noel Crespi
CNRS UMR5157, Telecom SudParis
Institut Polytechnique de Paris
Evry, France
noel.crespi@it-sudparis.eu

Abstract— In the digital revolution, even secure communication between individuals, services and devices through centralized digital entities presents considerable risks. Service providers collect and store information that is used for data mining, profiling and exploitation without users’ knowledge or consent. Having service providers continue to offer their centric solutions is inefficient in terms of duplication, has serious security lacunae and is cumbersome to the users. The Self-sovereign Identity (SSI) concept, which includes the individual’s consolidated digital identity and verified attributes, enables the users of data to exert their ownership and gain insights from their data’s usage. The authentication and verification of digital identity is essential to achieve the privacy and security of distributed digital identities. However, the current literature still lacks the comprehensive study on components of identity management as well as user privacy and data protection mechanisms in identity management architecture. In this paper, we provide a coherent view of the central concepts of SSI, including the components of identity proofing and authentication solutions for different SSI solutions. Firstly, we discussed an overview of Identity management approaches, introducing an architecture overview as well as the relevant actors in such a system and blockchain technology as solution for distributed user-centric identity. Then we analyzed the authentication and verification mechanisms in the context of digital identity. Finally, we discuss the existing solutions and point out the research gaps and elaborate challenges and trade-offs towards building a complete identity management system (IdMs).

Keywords: Blockchain, Self-sovereign Identity, authentication mechanism, Identity proofing, claim verification

I. INTRODUCTION

Recently, the rising surveillance and security breaches concern the user’s privacy in the current identity management ecosystem. To provide user-centric services, organizations assemble huge amount of personal information. The collected data is further utilized for profiling, prediction and economic growth. In many cases, the user has no insights about stored data about them and how it is used by service providers. The management of identities and personal identifies information (PII) is controlled by central authorities and user has little or no control over their data sharing and privacy. Furthermore, the collection of PII makes the service providers primary target of attacks and results in security breaches and privacy exploitation [1]. Many organizations have developed their authentication mechanism based on the OAuth protocol but central authority control remains intact.

The digital identity authentication ensures that individuals are who they claim to be in the online systems. The verification of subject and protection of sensitive information is the key component of trustworthiness in the identity management. Users have to exchange their personal information (e.g. credentials, PII etc.) with organizations in exchange of services. To overcome stealing, misusing or manipulating these data in central approach, services providers are required to provide many factor authentications along with management of identities which further complicates the systems [2]. Besides central approach, federated instances provides access to multiple sites with same credentials. However, the control and ownership of data still remains in the hand of identity service provider.

The contemporary approaches in identity management challenge the designers for expert security reviews and usability analyses concerning user experience and interaction with active agents in the identity infrastructure [3]. The recent work to eliminate the central service providers is one unique digital identity that is build, managed and controlled by identity owner [4]. Such identity that provides user centric data ownership is called self-sovereign identity (SSI). The blockchain technology recognized as temper resistant and transparent ledger [5]. Thus, it can be used to bind the users with the claims they make to prevent the identity frauds. Besides challenges of ownership, interoperability and controllability in SSI, the challenge of trustworthiness by management of evidences of digital identities demands prime attention. There is significant need for protection of digital identities through standardized solutions and interfaces for identity proofing and evidence exchange between subject, claim verifier and issuer. We have formulated our discussion
around the different notions of central topics in the identity management systems and blockchain technology disruption such as:

- The need to use Blockchain for identity management
- Authentication and User data Privacy
- An analysis of the existing solutions based on SSI architecture

Section II presents an overview of the identity management approach and describes the implications of managing identity using blockchain. An overview of the identity proofing components and authentication mechanisms for digital identity is explained in section III. The related research works are presented in section IV. Section V concludes the paper with summary of the remaining issues and main challenges for privacy and personal data protection.

II. BACKGROUND

A. Identity management approach

Identity management is an administrative process to create and maintain user account to be used for authentication and identification in online services. It is required to simplify the user provision process and ensure the rightful users can have access to the services. The identity management system (IdMs) life cycle comprises of four phases including enrollment, authentication, issuance and verification. enrollment and their role have been described in Table 1. We have quoted three types of IdMs with or without DLT according to our interpretation.

1) Centralized Identity

The central service provider remains the central power in this approach by collecting the user’s credentials and validate them to access the online services through their own authentication mechanism [6]. In DLT bases approach users’ credentials are validated through the central authority and further validation is process through identity information stored in DLT layer.

2) Federated Identity

The federated service provider separates the enrollment entity and other entities that rely on authentication process to verify the digital identity. An Identity provider, responsible for creating, maintaining and authenticating all users, plays the role of central hub for Various Web-based service providers [7]. User can enroll to service provider service A and can use the same identity to access the service b or any number of services allowed to be accessible through authentication process that validates the user’s claim [8]. The Facebook and Google single sign on are examples of federated entities.

3) Self-Sovereign Identity

The self-sovereign Identity (SSI) provides the ownership of data to user to promote user control and transparency. Based on rules of need-to-know and need-to-retain, the owner of data can control the information without relying on third parties that can result in data lost or misuse of sensitive information. “The user’s rights increased transparency with regulations, privacy by design systems, data portability and security”. The importance of self-sovereign based Identity systems has skyrocketed ascribable to current identity crises. Table 1 shows the actors involved in self-sovereign identity.

B. Blockchain Technology

The concept of blockchain first introduced as Bitcoin [9] is a peer-to-peer network and provide transparency through reaching consensus on transactions. The immutability of blockchain and consensus role eliminate the role of central authorities and appear to be ideal solution for distributed environment. As the data is the most valuable asset today, the implication blockchain in data driven architecture can bring features of decentralization, anonymity, audibility and persistency [10]. The most frequently mentioned terms of blockchain technology are described below:

i) Node and Block: Node is computer in peer-to-peer network representing the owner of transactions carried by certain user. Block is an immutable page of distributed ledger in the blockchain. After reaching consent on transaction, a block is added in the blockchain.

ii) Consensus: The consensus mechanism is used to process and validate the transaction through approval of decision of nodes. Widely used consensus algorithms include proof-of-work, proof-of-stake and Practical Byzantine Fault Tolerance.

iii) Scalability: In currently available solutions, the node scalability or performance scalability is provided depending upon access. Public blockchains such as Ethereum [11] and Bitcoin provide node scalability and Hyperledger [12] as private blockchain offers performance scalability.

iv) Smart contract: The third-generation revolution of blockchain broaden the application of blockchain in various domain besides asset management and cryptocurrency [13]. The complex applications can be controlled by smart contract by defining arbitrary rules. The functions in Ethereum smart contract have “gas” cost depending upon computational steps and storage space. The gas cost is paid in cryptocurrency called ether.

v) Access: Depending on consensus, the blockchain is categorized into three different types. Public or

<table>
<thead>
<tr>
<th>Actors</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentication Provider</td>
<td>Enrollment and</td>
</tr>
<tr>
<td></td>
<td>authentication of user for</td>
</tr>
<tr>
<td></td>
<td>rightful access to services</td>
</tr>
<tr>
<td>Attribute Provider</td>
<td>Provide selective disclosure</td>
</tr>
<tr>
<td></td>
<td>information for validation of</td>
</tr>
<tr>
<td></td>
<td>digital identity by identity</td>
</tr>
<tr>
<td></td>
<td>provider</td>
</tr>
<tr>
<td>Service Provider</td>
<td>Request claim verification by</td>
</tr>
<tr>
<td></td>
<td>issuance of credentials from</td>
</tr>
<tr>
<td></td>
<td>identity provider</td>
</tr>
<tr>
<td>Identity Provider</td>
<td>Validate PII and NPII to</td>
</tr>
<tr>
<td></td>
<td>present subject for what they</td>
</tr>
<tr>
<td></td>
<td>claim to be</td>
</tr>
</tbody>
</table>

| TABLE 1. IDENTIY MANAGEMENT SYSTEM AGENTS |
permission-less [14] blockchain provides the anonymity feature but lacks the privacy. While the private and consortium blockchain is used at organization level.

C. Motivation: Need of Blockchain

By leveraging the blockchain, the ten commandments of self-governed identity [15] can be achieved to overcome digital identity issues. The consensus mechanism fulfills the need of trustworthiness for verified published attributes. As blockchain is temper-resistant ledger, the persistent of claims can be attained. The SSI approach is user-centric and demands full control of user over their own data. Full control is given in the chain structure that provides chain per identity such as Trust chain [16] or The Tangle [17]. Such chain structure can also establish the existence principle for user to give the right to be forgotten. The claim Blocks withhold the personal information and authentication of claims ensures the protection and data minimization. The blocks can be shared with other platforms thus providing interoperability and portability.

III. BLOCKCHAIN FOR IDENTITY AND DATA MANAGEMENT

From security perspective, solutions and regulations are developed and already concerning personal data. The exchange of information between communication agents is ambiguous and hard to keep track of what data is shared as compare to actual granted access. The anonymity of identity is affected by degree of link-ability of personal data [18]. It is important to provide selective disclosure of PII and track PII to overcome issues of personal data privacy. As mentioned in [19], PII is defined as subset of information sufficient to identify the identity holder within set of subjects such as driving license, address, passport, name, date of birth etc. Furthermore, PPII is the subset of all the attributes of complete identity, where complete identity is the union of all attribute values for instance bank name, part of email, religion, partial name etc. [20].

As digital identity is compartmentalized into different context according to personal information, such as PII, potential PII and non-PII. The situation and role dominate the activation of identity properties for identification and user authentication in various settings. Many organizations have developed their own propriety authentication mechanism based on OAuth protocol. The regulation from standard bodies are currently practiced for data privacy and management around the globe. General Data Protection Regulation (GDPR) [21] is enforced by European Union to protect consumers by returning control of their identity data back to identity holders. GDPR complaint digital rights with SSI principle of user-centric identity in recognition to the need for an individual to manage and control his or her own data. It includes right of access, consent, data minimization, portability and existence (right to be forgotten).

A. User Data Privacy

The leakage of sensitive information representing identity can be destructive for both individuals and organization. To provide privacy in IdMs, personal identifiable information should imply no link-ability in revealing identities(s). The right choice of pseudonyms and authorization can facilitate the establishment of privacy enhancing identity management. Such privacy enhancing measures including minimalization of personal information. It ensures data presentation for need to know and need to retain scenarios.

The research in the domain of data privacy targeting the data anonymity has provided various solutions to protect personal identifiable information. The k-anonymity [22] approach provide solution of linking of information and requires the information set to be indistinguishable from other k-1 information records. Other solutions proposed diverse representation and distribution of sensitive data to ensure secure data sharing and minimalization of data disclosure in [23,24]. The other techniques such as perturbing data or encrypted information also have been used but inefficient for large scale distribution. Blockchain technology provides public verifiable open ledger of transactions and provide. In [25], the sensitive information privacy solution in IdMs using blockchain has been proposed. The proposed framework ensures the data ownership, interoperability and full access control by utilizing the blockchain for public identification and personal information of user is stored locally. However, the PII claim verification has not been considered in this solution. The existing blockchain based solutions have either utilized the permissioned blockchain access to ensure privacy or authentication is bind to decentralized identifier (DID) and sensitive information is stored on user’s device.

B. Identity Proofing and Attribute Assurance

The one function that is fundamental to IdMs is to distinguish one subject from another. In current identity management infrastructure, service providers need to identify users through claim verification of identity or certain attributes of the user. This process involves four actors: service provider, identity provider, attribute provider and claim issuer. As shown in the Fig 1., the subject presents the identity evidence to registration authority that inspect the identifiable information by identification information and provided identity information. The identity proofing often relies on various attributes (PII) ID, driving license, financial or telecommunication account, address etc. Third party agent that demands the identity proof interacts with the IPS also called claim verifier and proof is provide to relying party based on trust relationship. It is possible that user initiate the registration with identity provider without prior identity proofing and getting into identity proofing process upon concrete demand. As service providers can have varying requirements for identity attributes and their quality. In this case, single trust level to an identity provider is not sufficient.
to assign. Additionally, trust levels for attributes are needed to reflect the assurance quality of different digital identities hold by an identity provider [26].

In the claim verification process, selective presentation of verification factors is essential to data protection. The subject and the verifier have to negotiate which verification factors will be used in each particular presentation and allows the subject to choose variable kinds of verification factors (e.g. credential, biometric sample and private key or PII etc.) to present to a verifier. As mentioned in [27], two types of approaches have been discussed for claim verification. In first approach named identity registry model, claims are stored offline and with user consent provided to claim issuer for attestation. Second model called claim registry model which is extension of former one, holds the identifier in the blockchain as well as records of attested claims. A claim shall include validity periods, associated identity, meta-data information and algorithm for signature/encryption and holds one to one relation with claim-issuer. A framework is proposed for [28] trust information at the identity holder as assertion as well as attribute level according to Identity, Credential & Access Management (ICAM) standard. As a part of IdMs VeryIDX [29], the author in [30] also consider trust aspects of attributes more diverse and proposed to have different levels for the correctness of information subject-to-identity mapping.

By leveraging blockchain technology, IdMs can utilize user attribute-based credentials approach to implement identity and attribute validation protocol. The identity attributes can be stored offline and hashes of certificate are stored in the blockchain upon successful verification between claim-issuer and relaying party [31]. In [32], attribute certification solution for SSI has been provided using blockchain. The attribute certificate requests and certified attributes are centrally stored in a permissioned blockchain and linked to unique user pseudonym and managed through user wallet.

C. Decentralized Authentication

Public key Infrastructure (PKI) is responsible for public key management and authentication of correct mapping between users and their respective keys. There are two approach for authentication under public key infrastructure in practice. In centralized approach, the hierarchically structured central certificate authorities manage the certificate and holds the power to issue or revoke the certificate any time to keep secure authentication streamline [33]. On the contrary, decentralized authentication users can designate other parties as trustworthy to sign their certificates. This social trust mechanism is called PGP Web of Trust (WOT) [34]. In the process, the relying party can verify the person through provided certificate while withhold individual’s signature. While centralized structure makes CAs vulnerable, the authenticity of public key information leads to verification of malicious users’ keys that impose serious threat to system.

Recently various blockchain based PKI solutions have been introduced to overcome these issues. In [35], the bitcoin-based PKI authentication is built on top of Namecoin named as certcoin. The proposed solution provides the registration, update, revocation mechanism by creating two version of key pairs. In case of loss of online, key, the offline key is used to create new pair by verification of signature as old same key. For distributed authentication (verification and lookup), certcoin has exploited the kademlia DHT [36] for self-sustaining key-distributor service. Since, the increasing size of bitcoin blockchain requires huge storage space, the optimization techniques of cryptographic accumulators are employed to maintain content size. However, the empirical results and viability of proposed solution demands the practical implementation to uphold the scalability and optimization of solution as authenticating constrained devices using the Merkle Hash Tree accumulator is costly. Similar blockchain based WOT solution to replace centralized authorities and PGP WOT has been proposed in [37] named “Authcoin”. It focused on authentication and validation of public key to prevents attacks and provide detection of malicious keys. The validation which extends to public/private key, domains and accounts is handled through challenge-response procedure. Similarly, the authentication is followed in which involved parties verify in by taking challenge to verify identity such as send picture holding identity card etc. PGP approach for revocation of keys and signature is utilized. In [38], PKI called Claim-Chain based on decentralized key distribution system called is proposed. For authentication, claims by users are stored in blocks and each block hold reference of previous one to both audit past states of the chain, and authenticate the validity of newer blocks. In [39], JSON Web Token (JWT)-based authentication scheme leveraging blockchain has been introduced. The solution provides the registration of online identity using blockchain and protect personal cloud space API endpoint using JWT.

The bidirectional approach of verification and authentication is claimed to be more secure to detect and identify the sybil nodes. The use of existing domain authentication mechanism exposes this solution to malicious attacks which can be improvised. In [40], the blockchain based authentication system has been proposed to enforce access control. It is designed to provide privacy and security by integration cryptographic mechanisms such as attribute signature, authentication code and multi-receivers encryption. The blockchain based virtual zones in the distributed environment called bubble of trust is proposed in [41]. This solution is designed to provide the robust authentication and identification of devices and protect the availability and integrity of data. In [42], the allowed signature and public key are added within smart contracts for implicit authentication. The decentralized Ethereum blockchain based authentication mechanism has been proposed in [43]. The node is authorized in all the network clusters if it holds verified existence in one cluster.

IV. OVERVIEW OF IDENTITY MANAGEMENT SOLUTIONS

Many solutions have been proposed and developed from perspective of digital identity management and personal data security and privacy. We limit our discussion to the systems and architectures that proposed identity management and personal data using blockchains. There are no definitive
The representation of persistent identifier for a user identity through smart contract Id or blockchain Id

The identity (e.g. private key) is either stored on the device the identity created or utilizing the smart phone for key storage.

claim-> signature, name, validity period, scheme

Claims are stored offline
Public identifier stored on blockchain

Authentication is separated from authorization allowing others to change to also change the DID (e.g. quorum based key recovery, revocation protocol)

One-to one relationship between claim and claim-verifier

Central intermediaries (e.g. IPFS, centralized storage provider, local storage etc.) are used for personal identifiable information storage.

-Identity-attribute mapping (Ethereum claim registry, PGP and WOT, uPort registry model etc.)
- ShoCard attribute proofing

Sovrin, uPort, block stack, shocard, selfkey

TABLE 2: OVERVIEW OF SSI COMPONENTS IN EXISTING BLOCKCHAIN BASED IdMs

<table>
<thead>
<tr>
<th>Registration/Identification</th>
<th>Authentication</th>
<th>Identity Proofing</th>
<th>Personal Data Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>The representation of persistent identifier for a user identity through smart contract Id or blockchain Id</td>
<td>The identity (e.g. private key) is either stored on the device the identity created or utilizing the smart phone for key storage.</td>
<td>claim-&gt; signature, name, validity period, scheme</td>
<td>Claims are stored offline</td>
</tr>
<tr>
<td>Naming layer map to non-human readable system Id using Naming service (i.e. ENS) or DIF universal resolver</td>
<td>Authentication is separated from authorization allowing others to change to also change the DID (e.g. quorum based key recovery, revocation protocol)</td>
<td>One-to one relationship between claim and claim-verifier</td>
<td>Central intermediaries (e.g. IPFS, centralized storage provider, local storage etc.) are used for personal identifiable information storage.</td>
</tr>
<tr>
<td>IdMs: uPort, Blockstack [52], SelfKey, Civic</td>
<td>uPort, Sovrin</td>
<td>-Identity-attribute mapping (Ethereum claim registry, PGP and WOT, uPort registry model etc.) - ShoCard attribute proofing</td>
<td>Sovrin, uPort, block stack, shocard, selfkey</td>
</tr>
</tbody>
</table>

evaluation criteria to compare currently available identity management solutions.

However, they have been analyzed and compared under Identity Laws and SSI principals. In Table 2, the use of blockchain in implementation of SSI component based on currently available solutions have been shown. In [15], the architecture of famous blockchain based IdMs systems are discussed. In [44], The study of 23 identity management solutions have been discussed but does not provide conclusive comparison of solutions. While the comparative study of blockchain and non-blockchain identity systems is provided for 11 existing solutions in [45]. The evaluation is performed based on SSI principles.

In our discussion we have provided the overview of existing solutions in terms of personal data management and self-sovereign architecture. We have excluded non-blockchain based solutions in our analysis. In [46], uPort provide framework for users to gather attributes from an ecosystem of trust providers but does not provide identity proofing. For revocation in case of key lost, Quorum of blockchain is used. The smart contract id on Ethereum blockchain is used to represent public identity of user. It provides data ownership and selective disclosure however the privacy of user information in JSON data structure on message server can be compromised.

The, ShoCard [47] provides verification of identity is provided for online interaction and use store encrypted version of attribute certificate on server as backup. Use central server as intermediate between user and relying parties. The minimalization of data is not supported well. uses Bitcoin to record a commitment to personal data that was verified during identity proofing, and store the hashes of certifications that build upon the user’s Seal created by relying parties The Sovrin identity system facilitate with Identification on permissioned ledger.

The Sovrin IDM [48] use of attribute-based credentials that allows users to only reveal credentials that they choose with relying parties and WOT helps protect user against deception. For recovery mechanism, it relays on attribute-based shredding. It does not provide verification of relying parties so user needs to relay on WOT. User has full control over their identity but personal data protection is less secure as it lacks support for claim verification support. In [29], the claim verification blocks are designed in proposed IdMs to record the claims upon verification and maintenance. EverID [49] provides personal data ownership in hands of identity-owner and provide control of data sharing and data usage. Self-key [50] and LifeID [51] and Identity systems additionally fulfill provable property of identity system which means the claims of user identity and identity attributes can be verified by collecting information using zero knowledge proofs. The Blockstack [52] attempts to redesign the naming system in order to provide elucidation of Identity. It has PKI authentication features using state machines and storage aspect in blockchain to preserve privacy and resource identification.

V. RESEARCH CHALLENGES AND TRADE-OFFS

So far, we have explored the origin of digital identity and how blockchain can leverage the current rise of SSI. Nevertheless, there are still some challenges and trade-offs in building a feasible and effective IdMs. In this section, we point out few future challenges as follows:

- **Elimination of Intermediaries**: Each blockchain-based IDM solution offers the decentralized solution to
alleviate the control of centralized authorities. However, most of these solutions rely on central server or intermediaries for data storage and key revocation. The complete removal of CA can compromise several functions of identity management such as backup of cryptographic keys, identity recovery, lookup services etc. The poor management of identity by users in fully user-controlled system can affect the validity and data flow in infrastructure.

- **Privacy-enhancing identity management**: As the user-controlled identity demands transparent flow of data, identity management infrastructure should be designed to support pseudonymity while maintaining required degrees of confidentiality, integrity, authenticity, non-repudiation and robustness. In an authorized access to certain service as an anonymous user, identity holder needs to show authorizations to the service which are issued by a third party while they remain unlinkable to the users’ pseudonyms. The verification of third parties and mechanism to build trust relationship between service provider and third parties needs secure communication channel.

- **Scalability and Optimization**: Currently, many proposed novel solutions for IdMs using blockchain are either prototypes or developed systems with promised scalability in future research. The scalability and optimization aspects for distributed IdMs is essential to uptake adoption.

- **Communication parties trust**: The trust and reputation between attestation verifiers and relying parties is essential to certify identity attributes in WOT where any node can voucher other and difficult to quantify the trust anchors in the network.

- **User experience**: User experience and integrated human interaction is another challenge needs to be addressed by DLT based IdMs. The wide adoption of federated identity management by users suggest that novel solutions for identity management built upon same user interaction unlikely to uptake [51]. The authentication and identity proofing methods merely rely on unique identification through blockchain identifier and poor management of private keys by users limit the scope of IdMs for non-technical users.

**VI. CONCLUSION**

We have discussed the self-sovereign identity architecture and various blockchain based identity solutions that claims to fulfill the self-sovereignty. Blockchain technology can disrupt the traditional approach of identity management and this intervention can benefit the self-sovereignty vision of digital identity. As the identity is linked to personal information of user, the privacy and security risks for PII can compromise information in data breaches. The distinction between various subjects requires identity proofing and additional trust level for identity attributes. The decentralized authentication approach based on PKI can alleviate such risks by ensuring privacy and security through integration of cryptographic mechanisms.

We highlighted various solutions exist in order to solve the current identity management issues. Very recent solutions have been discussed in this paper. The ongoing research in this domain includes implementation along with proof of concept of proposed solutions-based principle of self-sovereign identity that now act as evaluation criteria. However, user empowering identity objectives such as giving full control in user’s hand raise potential privacy and controllability issues. These concerns are much needed to be addressed in future novel solutions or improvements in existing solutions.

**REFERENCES**


