Industry-Grade Self-Sovereign Identity

MSc Thesis

Rowdy Chotkan
R.M.Chotkan@student.tudelft.nl

Introduction

This document describes research into the development of an *Industry-Grade Self-Sovereign Identity* (IG-SSI) scheme. This scheme will be developed with collaboration of the Dutch Ministry of the Interior and Kingdom Relations and will serve as research into a digital identity scheme for the European Union. As this thesis is written per requirements of the 4TU Cyber Security programme, it will focus on applicable Cyber Security concepts and as such privacy and security will be the core of the design.

Problem Statement

As described by Allen (2016), there exist four different phases in identity management systems. Next, we describe each of them.

Phase One: Centralised Identity

With the onset of the Internet, centralised authorities such as IANA and ICANN became the issuers and authenticators of digital identities. E.g., the IANA determined the validity of IP addresses. Next, in order to generate trust through certificates, Certificate Authorities were created, which were able to also delegate some power through hierarchies. Finally, as mentioned by Cameron (2005), the distributed nature of the internet let to each platform implementing its own digital identity management in the form of e.g. user accounts. All of the above properties of the current Internet ecosystem are inherently centralised authorities. With the consequence of the user not owning any of his digital identities, as their are all either assigned to her or managed by others. Already in 1991, Zimmermann (1999) showed that distributed identity management is indeed possible, to some extend. However, Zimmermann's PGP was never widely adopted.

Phase Two: Federated Identity

The second generation of attempted to overcome the hierarchies, by imagining a *federated identity*. An example of this is Microsoft's Passport initiative, allowing identities across different domains, in this case, multiple websites. However, this initiative soon proved to be far from optimal, as it makes Microsoft the main authority. This was improved upon by allowing each site to remain an authority.

Phase Three: User-Centric Identity

The third generation attempts to put the user at the center of the identity. Examples of these include OpenID[1], OAuth[2] and FIDO[3]. The main goal of these implementation can be said to be user consent and interoperability, as the user has to provide consent for signing in on another domain using the methodology and they can be supported by any domain. However, the main drawback to these solutions are that the registering authorities can withdraw the digital identity at any time and, as such, there is still much to desire for user control.

Phase Four: Self-Sovereign Identity

The above limitations and designs failed to put the control in the user's hands. SSI aims to bridge this gap, by fully decentralising digital identities to such an extent that the user is in full control on what data is stored, what happens with said data, and with whom said data is shared.

Properties

As no consensus on a formal definition of Self-Sovereign Identity has been reached, the properties of SSI are loosely defined. There are, however, there are returning concepts in (academic) literature and common notions of use-cases. This section will aid in defining a set of requirements based on identified common themes in literature and will bridge the gap in unresolved issues.

One of the foremost motivation behind SSI, is its ability to generate trust in cyberspace. As presented by Cameron (2005), the Internet was built without an identity layer: there is no standardisation for authentication, authorisation and identification. As a consequence, the Internet consists of numerous workarounds of identification, which, evidently, has grown into a oligopoly of identity management held by large organisation such as Google, Apple, and Microsoft. The drawbacks of the current construction are quite broad:

Firstly, the data behind the identification measures, are not in the hands of the users. As a consequence, a user must ask permission to alter his data, has no direct access to his data, and has no control over how his data is processed. As these

[1] For *OpenID*, see `https://openid.net/connect/`

[2] For *OAuth*, see `https://oauth.net/`

[3] For *FIDO*, see `https://fidoalliance.org/`

identities are managed by commercial parties they are often prone to being processed and mined for the gain of said parties. Secondly, as these large organisation are no governmental entities, the resulting identities can never be used for legal identification purposes, an inherent shortcoming of their design. Finally, apart from the trivial overhead in different identification “workarounds", the lack of open-standards and centralised storage often leave such credentials in peril. As often no proper security requirements are set in place (e.g., a simple password), the credentials can either be easily brute-forced or stolen, resulting in identity loss. Their often centralised nature, can be weakness as well, as a security breach may impact the digital identities of all users.

The foremost common theme which can be said to have reached consensus, is the user-centric approach of SSI. Namely, the rationale of SSI's existence is making the user the manager of his own identity.

The most commonly discussed set of properties is that posed by Allen (2016). Allen posed the following set of *principles*, which are to ensure the user-centric nature of SSI. These consist of the following

1. **Existence**: users must have an independent existence. I.e., a (digital) sovereign identity does not solely exist digitally. As a result, it can be interpreted as requiring to be tied to a physical entity.

2. **Control**: users must have control over their identities. This entails a full authority over the user's own identity: the ability to share, update, and even hide.

3. **Access**: users must have access to their own data. Similarly to the above principle, users must be able to access the all of their own data.

4. **Transparency**: all involving systems and algorithms must be transparent. This entails open-standards and open-source software.

5. **Persistence**: identities must be long-lived. Identities should, thus, exists until destroyed by the user.

6. **Portability**: information and services about identity must be transportable. I.e., identities must not be held by a single third-party, as they may not support it livelong. This principle would be satisfied by the *Control* and *Persistence* principles.

7. **Interoperability**: identities must be as widely usable as possible. This ensures that the identities can be globally deployed and can be achieved partly by adopting the *Transparency* principle.

8. **Consent**: users must agree to the use of their identity. This principle strengthens the *Control* principle, as sharing of attributes may only occur with the consent of the user. However, the Allen noted that this must not require interactivity.

9. **Minimalisation**: disclose of claims must be minimised. I.e., the minimal amount of information must be disclosed when sharing claims. This principle is focused on privacy and prevents misuse of data.

10. **Protection**: the rights of users must be protected. The right of users must take precedence over the identity network itself. This can be achieved thorough the *Transparency* principle and decentralisation.

The above set of principles is often adhered to as a set of requirements. See e.g. . In addition to these ten principles, Add re
Stokkink and Pouwelse (2018) add the principle of *Provability*: claims must be provable, as otherwise they can be deemed worthless.Tobin and Reed (2016) build upon these ten principles by subdividing these into three categories:

- **Security**: aims to keep the digital identity information secure. This consists of: *Protection*, *Persistence*, and *Minimisation*

- *Controllability*: focuses on the user-centric foundation of SSI. This consists of: *Existence*, *Persistence*, *Control*, and *Consent.*

- *Portability*: this requirement results in the user not being tied to a single provider and being able to use their identity without bounds. This consist of: *Interoperability*, *Transparency*, and *Access*.

The additional principle defined by Stokkink and Pouwelse (2018) can be categorised into *Security*, as the provability of claims aids in generating trust and in authentication.

The work set out by Cameron (2005), is another commonly cited set of principles for SSI. In their work, Cameron developed the so-called *Laws of Identity*. These laws explain the shortcomings and successes of digital identity systems and, as such, are applicable to SSI. These consist of the following:

1. **User control and consent**: digital identity systems must only reveal personal identifiable information (PII) given prior consent by the user. Through this law, trust can be built between the system and the user.

2. **Minimal disclosure for a constrained use**: the solution which discloses the least amount of and best limits the use of PII, is the most stable long term solution. This law minimises risk, as it is assumed that a breach is always possible.

3. **Justifiable parties**: disclosure of data with third parties must always be justifiable in a given identity relationship. Through this law, the user is aware of any third parties with whom is interacted with whilst sharing information.

4. **Directed Identity**: universal digital identity systems must support “omni-directional" identifier, which can be said to be public, and “unidirectional" identifiers, which can be said to be private, enabling identification whilst facilitating privacy.

5. **Pluralism of operators and technologies**: universal identity system must support multiple identity technologies run by multiple identity providers. This law enables the incorporate this somewhere, disallowing vendor lock-in and encourages the use of open-standards.

nostic

6. **Human integration**: universal digital identity systems must incorporate the user as a component of the system, offering protection against identity attacks. This laws attempts to bridge the discontinuity between the actual (human) users and machines with which they communicate.

7. **Consistent experience across context**: universal digital identity systems must allow for a separations of domains, whilst enabling a consistent experiences within and across them. This law thus enables interoperability across different operators and technologies.

Related Works

Mühle, Grüner, Gayvoronskaya, and Meinel (2018)

Mühle et al. (2018) describe an overview of SSI. They state that ISS differentiates itself with traditional identity management systems by being a user centric model as opposed to service provider centric. They describe two architectures for SSI: the *Identifier Registry Model* and the *Claim Registry Model*. Wherein the former model the pairing of identifiers and public keys of users are stored onchain and claims offchain. In the later model, in addition to serving as a registry for identifiers and public keys, the claims themselves are also stored onchain. Next, they focus what they deem the four core components of SSI: identification, authentication, verifiable claims, and attribute storage. Identification comes done to the issue of having both uniqueness and human-readability in identifiers of clients. It is noted that the current best effort is that of *decentralised identified* (DID), which has a universal resolver by the Decentralized Identity Foundation[4].

Der, Jähnichen, and Sürmeli (2017)

Der et al. (2017) describe the o opportunities and challenges for a digital revolution caused by SSI. The authors start with explaining the terms *digital identities* and *secure digital identities*. Where a *digital identity* is a temporal reflection of a regular identity: it merely contains specific characteristics of an identity, with varying level of detail. A digital identity can be held by any type of entity, may it be a person, a car, or a device. It usually has to function to use a particular service. In addition, a *secure digital identity* adheres to the requirements of *privacy* and *trustworthiness*. Where privacy leads to only authorised access to the identity, and trustworthiness the correctness of the attributes contained in the digital identity.

The authors then explain the general concept of Self-Sovereign Identity. They state that SSI can be the next step in identity management and mention the ten principles by Allen (2016). SSI moves the requirements of privacy and trustworthiness to the user, requiring the user to provide evidence.

Next, three opportunities for SSI are explained. Firstly, SSI can counteract the oligopoly present in the management of current digital identities. Secondly, it can provide help to people living in crisis areas, as identities may no longer require ties to local government. Finally, SSI may help companies to adhere to the GDPR as privacy can be more easily implemented.

The challenges for SSI are also explained. It is stated that current digital identity services (e.g. Facebook connect) allow for a certain level of comfort by trading in a certain level of control of their identity. Based on that assumption, the case is made that one of the core challenges of SSI is that the additional required administrative efforts of SSI must be sufficiently comfortable. The following key challenges are outlined:

- Protection of privacy across transactions.
- Transparency between two parties during a transaction, i.e., consensus on content and conduct.
- Persistency of digital identities and logs for long-term transparency.
- Trustworthiness of digital identities and claims.
- Consistency between granted rights and real usage.
- Standardisation of data formations and interfaces.

Finally, the efforts by the ISÆN and an outlook are given with applications of SSI for the Internet of Things and institutions.

[4] `https://identity.foundation/`

Stokkink and Pouwelse (2018)

Stokkink and Pouwelse (2018) present a blockchain-based digital identity solution. It is stated to be an academically pure model for SSI. They state that the first half of the problem regarding the creation of such a model, is the need for Self-Sovereign Identity: identity holders must be identity owners. The second half of the problem is the need for legally valid signatures: identities can e.g. be recognised by the governments, making them legally valid. They firstly describe the solution for the first halve of the problem, in which they state the ten principles by Allen (2016). The blockchain-nature of their solution is said to intrinsically satisfy the majority of the principles, apart from:

- Portability
- Interoperability
- Minimalisation
- Protection
- Provability (added by authors)

Othman and Callahan (2018)

Othman and Callahan (2018) describe their Horcrux protocol, a decentralised biometric credential storage option via blockchain using W3C's Decentralised Identifiers (DID). The authors mention that the current drawback of traditional biometric-based authentication systems is that the systems are a single point of compromise for securing digital identities. This is caused by requiring a central authority for storing templates of biometric samples. The Horcrux protocol combines the SSI ecosystem with the h 2410-2017 IEEE Biometric Open Protocol Standard (BOPS). This is performed by dividing biometric templates into $n \leq 2$ shares, which are then stored distributed-wise. The actual shares are stored offchain, but resolvers to the DIDs are stored on onchain.

Ferdous, Chowdhury, and Alassafi (2019)

Ferdous et al. (2019) describe a mathematical model for SSI in order to provide a formal and rigorous treatment of the concept of SSI itself. As such, they firstly formalise a mathematical definition and identify the required properties for SSI, after which they investigate the impact SSI can have using the Laws of Identity. Finally, they investigate the implication of applying blockchain technology to SSI. Their formalised model of an SSI contains the definition of an entity. An entity has an identity which consists of of the union of all its partial identities. These partial identities are all of his attributes and values in a specific domain. Hence, an entity can be contained in multiple domains, where each partial identity can be subdivided into profiles (subsets of the attributes contained in the partial identity within a domain).

Cameron (2005)

Cameron (2005) describes one of the inherent flaws of the Internet being the lack of an identity layer: there is no standardised mechanism for identification, resulting in a shattered "patchwork of identity one-offs", so-called workarounds for identification.Cameron proposes a *unifying identity metasystem*, which, similarly to what sockets provide for networking, provides an abstraction for identification which allows application to abstain themselves from specific implementations and allow (lose) coupling of digital identities. For this, Cameron developed the seven *Laws of Identity*. These will be discussed more thoroughly in section .

Allen (2016)

Allen (2016) discusses the ten principles of SSI. Firstly, their work explains issues with traditional (physical) identity measures, e.g. driver licenses and social security cards, which are erroneously portrayed as identities. As a consequence, the issuing authority has the capability to nullify ones "identity".Allen propose SSI as an improvement and solution. Next, the four phases of evolution of identity are explained.

Contributions

The work set out by Stokkink and Pouwelse and Stokkink, Epema, and Pouwelse will serve as a foundation of the IG-SSI scheme. The contributions made by this thesis will be an SSI scheme that can be said to be of *industry-strength*, which will be substantiated with a real-life trial of an implementation of said scheme. The main knowledge gap currently existing in the research area of SSI is the gap between the theoretical frameworks and the feasibility of these theories. E.g., strict processing latency requirements on mobile devices, communication overhead, and fault-tolerance. As such, this thesis will attempt to bridge this gap by constructing an SSI scheme together with developing an interaction model that allows for a practical implementation that is to be verified through real-life user tests.

Research Questions

The topic of Self-Sovereign Identity and the notion of *Industry-Grade Self-Sovereign Identity* shall foremost be investigated through the following research question:

"How can Self-Sovereign Identity serve as a digital alternative to centralised identification measures?"

This research question will allow for the investigation into and the development of a state-of-the-art SSI architecture. Based on the identified knowledge gap, the following sub-questions can be investigated:

1. *How to store verifiable claims in a decentralised fashion?*
2. *How to integrate the concept of trusted entities (which allow for claim verification) into Self-Sovereign Identity?*
3. *How to design an open Self-Sovereign Identity standard that allows for an accessible implementation (e.g. supported by all major smartphone operating systems?)*
4. *How to integrate an open interface for secure hardware tokens?*
5. *How to integrate an open interface for (biometric) authentication technology?*

Based on these results, we will be able to design an SSI architecture that will overcome these shortcomings and be deemed to be of *industry-strength*.

References

Allen, C. (2016, 5). *The Path to Self-Sovereign Identity.* CoinDesk. Retrieved from `https://www.coindesk.com/path-self-sovereign-identity`

Cameron, K. (2005). The laws of identity. *Microsoft Corp*, *5*, 8–11.

Der, U., Jähnichen, S., & Sürmeli, J. (2017). Self-sovereign identity - opportunities and challenges for the digital revolution. *arXiv preprint arXiv:1712.01767*.

Ferdous, M. S., Chowdhury, F., & Alassafi, M. O. (2019). In search of self-sovereign identity leveraging blockchain technology. *IEEE Access*, *7*, 103059–103079.

Mühle, A., Grüner, A., Gayvoronskaya, T., & Meinel, C. (2018, 11). *A survey on essential components of a self-sovereign identity* (Vol. 30). Elsevier Ireland Ltd. doi: 10.1016/j.cosrev.2018.10.002

Othman, A., & Callahan, J. (2018, 10). The Horcrux Protocol: A Method for Decentralized Biometric-based Self-sovereign Identity. In *Proceedings of the international joint conference on neural networks* (Vol. 2018-July). Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/IJCNN.2018.8489316

Stokkink, Q., Epema, D., & Pouwelse, J. (2020). A Truly Self-Sovereign Identity System. *arXiv preprint arXiv:2007.00415*.

Stokkink, Q., & Pouwelse, J. (2018). Deployment of a blockchain-based self-sovereign identity. In *2018 ieee international conference on internet of things (ithings) and ieee green computing and communications (greencom) and ieee cyber, physical and social computing (cpscom) and ieee smart data (smartdata)* (pp. 1336–1342).

Tobin, A., & Reed, D. (2016). The inevitable rise of self-sovereign identity. *The Sovrin Foundation*, *29*(2016).

Zimmermann, P. (1999). *Why I Wrote PGP*. Retrieved from `https://www.philzimmermann.com/EN/essays/WhyIWrotePGP.html`