Industry-Grade Self-Sovereign Identity

Article

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

Introduction

SINCE the dawn of the Information age, digital trust has been an issue requiring many workarounds. The core concepts of the internet are simply not built with trust in mind; there exists no standardised identity layer (Cameron, 2005). As a result, the current landscape of identification and authentication mechanisms form a digital ecosystem of "digital one-offs" (Cameron, 2005). As a consequence, the popularity of these digital one-offs by early pioneers of the Internet has resulted in an oligopoly in digital identity of Big Tech companies. This oligopoly results in an asymmetrical control of digital identities held by Big Tech. Wherein a regular oligopoly, consumers are at a disadvantage price-wise (Stigler, 1964), in this technical oligopoly, these identity providers have an asymmetrical control of ones digital presence and the ability to nullify access to such services in case one violates their terms of service. In addition, this oligopoly results in large information asymmetries as Big Tech has increasing amounts of knowledge on their users.
Furthermore, increasing needs for digital identities from governments such as the European Union, has catapulted the research and relevancy of the field itself. This need stems mostly from urgency of COVID-19 vaccination passports, requiring digital verifiability and validity across borders. As a result, this digital and socio-economical gap can prove to be filled by the concept of *Self-Sovereign Identity* (SSI).

SSI aims to fill the gap in digital trust by providing verifiable digital identities, putting the user at the center. SSI is an issue requiring multiple state-of-the-art technologies to be realised, thus, the feasibility of developing a schema that is both technologically and usability-wise sound, can be proven to be hard. Numerous solutions exist (e.g. Sovrin[1] and Serto[2], however, many require proprietary technologies or hardware, or require specialised infrastructure limiting equality in the network. As SSI combines multiple technologies, such as decentralised ledger infrastructure, public key infrastructure, and secure data management, many of the existing solutions do not stem strictly from academia, making their results more difficult to reproduce and limiting the analysis of their design choices.

This article introduces *Industry-Grade Self-Sovereign Identity*: a purely academic Self-Sovereign Identity framework focusing on an open standard, with intrinsic equality across the network. The scheme is based on the previous works by Stokkink and Pouwelse (2018), Stokkink, Epema, and Pouwelse (2020) and builds upon the IPv8 protocol stack. **[TODO:** Needs citation**]**. The main contributions of this work are a functioning SSI scheme, which can be said to be of *industry-grade*. IG-SSI makes the following contributions to the work set out by Stokkink and Pouwelse (2018): (1) Trusted Authority (TA) concepts, (2) offline verification, (3) Hybrid Revocation Modal (HRM), and (4) an Android reference implementation. Client communication runs directly from client to client without the necessity of external infrastructure.

The Problem Statement

The Internet was created without an Identity Layer. As a consequence, there is no standardised methodology for creating trust digitally; each digital service requires a (custom) implementation for digital identity management in order to both identify and authenticate users. Cameron (2005) coins the Internet's effort of identity management systems a "patchwork of identity one-offs", as every Internet service requires an identification workaround. The consequences of these identification workarounds are both for service providers and end-users sub-optimal.

For service providers, identification measures can prove to be a double-edged sword: whilst it allows them to manage their users' digital identities, it can also prove to be a burden in case their systems are compromised. The leakage of Personal Identifiable Information (PII) cannot only lead to liability in accordance to the GDPR (The European Parliament and Council, 2016) (e.g., the GDPR has the possibility to fine companies in the millions), but can also have side effects on the users. For instance, in case passwords are compromised, other services utilised by the user may be at peril or the leaked PII can be used for spear-phishing attacks. On the other end, users suffer from the same consequences and more: firstly, users must keep track of all their fragmented digital identities, often requiring to manage a multitude of identification credentials. A report published by LastPass

[1] For Sovrin, see: `https://sovrin.org/`

[2] For Serto, see: `https://www.serto.id/`

in 2019, shows that on average employees of small businesses manage 85 passwords. With the statistic that the use of brute-forced or stolen credentials are responsible for over 80% of the vulnerabilities utilised in breaches (Verizon, 2020), credentials continue to be a weakness in online identification measures. Secondly, users' information is stored in numerous amounts of locations, significantly increasing the chances of their personal identifiable information to be stolen, as this increased the attack surface. For instance, Thales (2020) reported that in 2020, 49% of US companies reported a digital breach of some degree.

Whilst since the dawn of the Internet digital identities have gone through multiple phases, the current landscape has resulted in a oligopoly of Big Tech companies providing digital identity management services.

The Current Landscape

Currently, identity management systems are in, what Allen (2016) refers to as, the third phase. This third phase is called the "*User Centric Identity Management*", originally described by Jøsang and Pope (2005). **[TODO:** Allen refers to another source (?)**]** These systems focus on user centricity and interoperability, allowing users to select their own provider. However, these efforts still resulted in the register being the owner of the identity, instead of solely the user. The main drawback to the current phase, however, is the introduction of initiatives such as Facebook ConnectMorin (2008) (contemporary known as Facebook Login[3]) or Google Identity[4]. Whilst these initiatives do allow selective sharing of identity information and regard user consent, they still store identities in a centralised fashion and are managed by a single authority. The global adoption of these digital identity providers, has led to an oligopoly of digital identities. **[TODO:** needs citation**]**

The oligopoly poses additional threats to users. The main issues regarding this oligopoly are lack of control, privacy, and information asymmetries. More prominently, Big Tech now has the ability to potentially revoke ones digital identity without warning, resulting in a loss of access to possibly countless of services. Privacy is at peril as Big Tech is enabled to gain information on their users through other services. This privacy concern can lead to market mechanisms such as information asymmetries, due to these extra opportunities for data farming. These identity providers are essentially commercial parties, profiting from data received through managing these identities. This breach of privacy often comes hand in hand with the free to use service offered by the digital identity service providers. The often circulating quote "If you are not paying for it, you're not the customer; you're the product being sold"[5] holds up in this regard. The issue with commercially available identities is that they do not provide legally valid identities and pose a huge threat on privacy, as the subject has no control over with whom their data is shared. This additionally leads to information asymmetries: as these big-tech companies posses large amount of PII of their users, any economic transaction made with them, results in them possessing more knowledge than the buyer. This effect has been regarded by Tobin and Reed (2016) as the use of adhesion contracts, which go against the users' best interests. These concerns portray a need for a different approach to identification, breaking the oligopoly and creating the ability to generate trust over the Internet.

As a result, there exist two reasonings for the rationale of Self-Sovereign Identity's existence: the first reasoning is to devoid the aforementioned current oligopoly of big-tech companies in the digital identity domain.The second reasoning is economic inclusion: residents residing in countries devoid of proper (central) identity infrastructure, are excluded from essential services enabled through identification system. World Bank Group (2016) defines identification to be required for the following:

- Inclusion and access to essential services: e.g., healthcare, education, and financial services.
- Effective and efficient administration of public services, policy decisions and governance.
- Accurate measure of development progress in areas.

Hence, without any form of valid identification measures, these residents are devoid of essential services and are less likely to be able to improve their living conditions or receive aid.

As the first issue, mostly regarding privacy and control, is a far more relevant topic in Computer Science, with the second problem is more a socio-politic issue, the primary focus of this research will be targeted at combating the former phenomena.

Background Information

In order to satisfy the self-sovereignty aspect of SSI, several works have composed principles and laws which must be adhered to. The most commonly discussed set of properties is that posed by Allen (2016), these consist of the following ten principles:

[3] For *Facebook Login*, see: `https://developers.facebook.com/docs/facebook-login/`

[4] For *Google Identity*, see: `https://developers.google.com/identity`

[5] `https://www.metafilter.com/95152/Userdriven-discontent#32560467`

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.

In addition to these ten principles, Stokkink and Pouwelse (2018) add the principle of *Provability*: claims must be provable, as otherwise they can be deemed worthless. As becomes apparent from these principles, the users are the most important aspect in this system. They take precedence over the protocol itself. Integral to this, is that online identities are no longer separate entities or, what biometrics refer to as pseudo-identities (Delvaux et al., 2008).

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 technologically agnostic property **[TODO:** incorporate this somewhere], disallowing vendor lock-in and encourages the use of open-standards.

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.

Design

Revocation

Revocation is one of the main issues in Self-Sovereign Identity. As in real life contracts and other agreements may become invalid before their termination date, the ability to revoke attestation in SSI must be available as well. Several motivations exist for revocation:

- Erroneously signed data: in case data was signed accidentally.
- A Legally invalid contract: in case at a later instance it became apparent that the signed data can not be legally upheld.
- Premature termination of a contract: in case a certain breach of contract occurs.

Note that expiration is not one of these listed motivations, as time-bound attestations can be realised using signed metadata. It is important that revocation can never occur due to expiration, as some claims should never be able to be revoked. For instance, it should not be possible for an authority to revoke a signature indicating someone is of legal age (unless in the rare instance that it was erroneously signed and can be publicly verified that this was, indeed, the case), as this fact can never become false.

As IG-SSI is built without specialised validation nodes, present in some blockchain-based protocol such as Zhou, Li, and Zhao (2019), there is no trivial non-interactive solution of revocation of signatures. The trivial solution is to actively query signees (i.e., the responsible authorities) and verify that they still attest for the signed information. There exist multiple problems with this solution. Firstly, this querying requires interactivity with the signee(s) of an attestations. Whilst interactivity is not a problem per se, it does introduce additional overhead. Firstly, it requires the signee(s) to be online. Whilst availability often is a key characteristic in distributed systems, there is no guarantee that specific clients, i.e. the signees, are available. Secondly, this interactivity generates additional overhead in the verification process. Apart from challenging the presenting client, the signees have to be actively queried, introducing additional verification time and network traffic. Secondly, as a requirement for enabling this interactivity, a (network) connection to the signees must be available. This completely nullifies the possibility for offline verification. Next, we discuss our solution for revocation: *The Hybrid Revocation Model* (HRM). This model requires no additional interactivity during verification and enables offline-verification.

Hybrid Revocation Model

The hybrid revocation model attempts to overcome the hurdle of interactivity whilst allowing for flexibility, enabling offline-verification. As IG-SSI is fully distributed and as such, each node is equal. As a consequence, the client performing the verification must be aware of any revocations belonging to a presented attestation. Selecting specific nodes for distributing and holding revocation, would deteriorate the equality principle. As these nodes would, then, possess the ability to hide certain revocations from the network or from certain peers. In order to guarantee the validity of said revocations, each peer should posses the revocations. With this principle in mind, the Hybrid Revocation Model (HRM) was designed.

In HRM, each peer has the possibility to posses the same information about revocations. Revocations are propagated through the network, enabling each peer to store revocations from clients they trust. This concept builds upon the notion of trusted authorities: each client has the ability to trust certain authorities. These authorities are referred to as Trusted Authorities (TAs).

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.

Delvaux, N., Chabanne, H., Bringer, J., Kindarji, B., Lindeberg, P., Midgren, J., … Skepastianos, D. (2008). Pseudo identities based on fingerprint characteristics. In *Proceedings - 2008 4th international conference on intelligent information hiding and multimedia signal processing, iih-msp 2008* (pp. 1063–1068). doi: 10.1109/IIH-MSP.2008.327

Jøsang, A., & Pope, S. (2005). User Centric Identity Management. *AusCERT Conference 2005.* Retrieved from `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.60.1563&rep=rep1&type=pdf`

LastPass. (2019). *THE 3RD ANNUAL GLOBAL PASSWORD SECURITY REPORT* (Tech. Rep.). LastPass. Retrieved from `https://lp.logmeininc.com/rs/677-XNU-203/images/LastPass_State-of-the-Password-Report.pdf`

Morin, D. (2008, 5). *Announcing Facebook Connect.* Retrieved from `https://developers.facebook.com/blog/post/2008/05/09/announcing-facebook-connect/`

Stigler, G. J. (1964). A theory of oligopoly. *Journal of Political Economy*, *72*(1), 44–61. Retrieved from http://www.jstor.org/stable/1828791

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).

Thales. (2020). *2020 Thales Data Threat Report* (Tech. Rep.). Thales. Retrieved from https://cpl.thalesgroup.com/sites/default/files/content/research_reports_white_papers/field_document/2020-04/2020-data-threat-report.pdf

The European Parliament and Council. (2016). *Regulation (EU) 2016/679 of the european parliament and of the council.* Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679#d1e6226-1-1

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

Verizon. (2020). *2020 Data Breach Investigations Report* (Tech. Rep.). Verizon. Retrieved from https://enterprise.verizon.com/resources/reports/2020-data-breach-investigations-report.pdf

World Bank Group. (2016). *Identification for Development Strategic Framework* (Tech. Rep.). World Bank Group.

Zhou, T., Li, X., & Zhao, H. (2019). EverSSDI: Blockchain-based framework for verification, authorisation and recovery of self-sovereign identity using smart contracts. *International Journal of Computer Applications in Technology*, *60*(3), 281–295. doi: 10.1504/IJCAT.2019.100300