# Decentralizing Digital Identity: Open Challenges for Distributed Ledgers

2018 IEEE European Symposium on Security and Privacy Workshops

Paul Dunphy[1], Luke Garratt[1], Fabien Petitcolas[2]
Innovation Center, VASCO Data Security, {[1] Cambridge (UK), [2] Brussels}
{firstname.lastname@vasco.com}

***Abstract*—Distributed Ledger Technology (DLT) has been proposed as a new way to incorporate decentralization into a wide range of digital infrastructures. Applications of DLT to digital identity are increasing in prevalence, with a recent survey reporting that 55% of DLTs in development track digital identity. However, while proofs of concept, open source software, and new ideas are readily available, it is still unclear the extent to which DLT can play a role to underpin new forms of digital identity. In this position paper, we situate this fast-moving application domain into the broader challenges faced in digital identity, with the aim to highlight the socio-technical nature of the challenge at hand, and to propose directions for future research.**

## I. Introduction

Distributed Ledger Technology[1] (DLT) is being investigated as a technique to incorporate decentralization into digital infrastructures that exhibit limitations in terms of e.g., reliance upon costly intermediaries, unwanted centralisation, or lack of transparency. DLT is a crucial underpinning of cryptocurrencies, which are now a permanent fixture in public discourse on financial investments. The increasing public literacy with concepts of cryptocurrency have given rise to efforts to apply DLT to a wide range of everyday applications. Of these new applications, a recent study suggests that 55% of surveyed organizations working on DLT were designing to "track digital identities" [13].

'Digital identity' is a term that can be interpreted in different ways, but it is most simply understood through the *three-party model*, which captures interactions between three entities: an end-user, a relying party (or service provider) and an identity provider. Within a defined scope, the digital identity of a user is the set of all attributes and identifiers related to that user. But while progress has happened quickly and has generated no shortage of white papers and open source software, it is still unclear how successful the intervention of DLT into digital identity can possibly be. Partly because the discipline of digital identity is highly contextual, and its complexity has ushered into disuse numerous cryptographically advanced technologies, such as Passport, Infocard, uProve amongst others. Therefore, it is a pressing question as to which course research can take to explore the feasibility of DLT-based identity technologies to maximise potential uptake and avoid a similar end-state of technology disuse.

In this position paper, we build upon existing work [10] and situate this fast-moving application domain of DLT into the broader context of digital identity to highlight the socio-technical nature of the challenges at hand, and to propose pressing directions for future research.

## II. Brief Background

One framework that captures the processes for creating, managing and using digital identities is *Identity and Access Management (IAM)* which has four components: authentication, authorization, user management, and user directories. Centralized infrastructures are the most common, where identifiers and attributes are created, owned by, and only valid within, a single organization; federated architectures enable the sharing of identifiers and attributes amongst organizations that participate in a defined *circle of trust*. A significant development following federated identity is *user-centricity*, a concept well covered by Bhargav-Spantzel et al. [2]. User centricity focused on providing decentralization of identity and enhancing user privacy and *control* over identifiers and personal data. Interest in applying DLT to digital identity has evolved alongside Bitcoin [15]. Bitcoin provides no mechanism to underpin trust in identities but relies upon unique pseudonyms bootstrapped by cryptographic public keys; an idea first proposed by Chaum [6]. In relation to DLT, Dunphy and Petitcolas [10] provide a glimpse of how DLT has been applied to digital identity by evaluating three representative DLT-based schemes: uPort, ShoCard, and Sovrin. This work uncovered dominant design approaches; surfaced privacy concerns; and highlighted a lack of focus on the end-user experience.

## III. The Big Challenges

There are enduring problems related to digital identity that, we believe, any DLT-based approaches must aim to address, or at the very least, must not exacerbate:

- **Identity fraud** – $16 billion was stolen from 15.4 million U.S. consumers in 2016, compared with $15.3 billion and 13.1 million victims a year earlier [17].
- **Data breaches** – Recently, the largest data breach of the 21st century occurred at Equifax where 143 million identity data records were breached.
- **Lack of reusability of identities** – Creates costs for organisations, data replication, and usability challenges for users. One survey suggests financial institutions spend on average $60 million per year to onboard new customers, while, some spend up to $500 million [18].

[1] A decentralized append-only ledger maintained by participants on a peer-to-peer network.


DOI 10.1109/EuroSPW.2018.00016

## IV. A Research Agenda

Our early exploration of this domain has created an interest to determine the problems in digital identity that DLT-based identity schemes can -- and cannot -- address. Investigating this question requires a course of research that aims to: (i) *understand* how DLT and other decentralising techniques interact with pre-existing challenges in digital identity (e.g., trust, interoperability, deployability); (ii) *envision the potential* that decentralised infrastructures might bring to underpin new approaches to digital identity. In this section, we share our thoughts on a starting point for applied research by proposing pressing areas that are worthy recipients of discourse and investigation.

### A. Refine understanding of DLT properties leveraged for identity

Mainstream discourse typically hints that DLT provides unquestioned benefits of transparency, immutability, auditability, and decentralization. However, it is important to critically question how these properties are born out in digital identity, and be able to recognize how these properties are prioritized in the design of existing schemes; lest we be unable to meaningfully distinguish and compare identity schemes that leverage DLT. As an example, Table I illustrates some known DLT-based identity approaches and an initial proposal of their dominant design priorities. We have only picked few DLT-based-approaches and related techniques, that serve as key exemplars of prevalent design decisions found in numerous other schemes and that provided sufficient technical details about their functioning.

*Transparency:* digital identity has a tense relationship with transparency because of the importance of privacy. On the one hand, transparency about the processes and procedures used by identity providers is crucial, and required by certain legislations (e.g., GDPR). On the other hand, transparency of information is not always desirable as the Swedish open data policy has shown [12]. If transparency is a design goal, research has shown that this can be achieved without DLTs. Certificate transparency (CT) [19] is an augmentation of the public key infrastructure (PKI) used for web domains that gives visibility to all certificates generated by a certificate authority. However, rather than using a DLT, CT puts to work an ecosystem of intermediaries with different roles i.e. *logs*, *monitors* and *auditors*. The logs maintain an append-only list of certificate records, while the role of the latter two entities is to validate the logs. Keybase.io uses a similar idea to register a person's different social media personas. Chase and Meiklejohn [5] define an abstraction of certificate transparency as a *transparency overlay* that can be applied to different problems. But how can society benefit from transparency over digital identities without compromising privacy?

*Immutability:* it should be difficult for a single or small group of entities to illegitimately alter historical data that group consensus has added to a ledger. Different DLT-based identity schemes prioritize different data items to be stored in an immutable ledger. For example, uPort allows identity claims to be stored in the Ethereum ledger. So why should claims be embedded in an append-only ledger, potentially long after they are relevant? One reason why this design decision could have been taken is to facilitate *credential-focused* [2] user-centric identity and may have chosen to provide the user with a means of safekeeping for those credentials – in this case on a public permissionless DLT. But of course, there is a trade-off between leveraging immutability, the resulting transparency of a publicly viewable ledger, and privacy. Comparing schemes such as uPort and Sovrin suggests a property of *auditability* can be achieved without placing identity claims in an immutable ledger. Stricter privacy laws such as GDPR give users increased rights of rectification and erasure of personal data. This reinforces the need for an intuitive separation between identities and claims and suggests differences in how each should be stored. So, when designing an ecosystem of DLT-based digital identity, how can we best leverage the qualities of immutability, and when should it be avoided in order to respect privacy?

TABLE I. Proposed design priorities of representative DLT-based Identity schemes. Certificate Transparency included for comparison since while it is not a DLT it achieves a DLT-relevant goal (transparency).

| Scheme | Primary | Secondary |
|---|---|---|
| uPort | Immutability | Transparency |
| Sovrin | Decentralisation | Auditability |
| ShoCard | Immutability | Auditability |
| Keybase.io | Auditabiity | Immutability |
| Certificate Transparency | Transparency | Auditability |

*Decentralization:* by creating decentralization in a typically centralized system, the goal is to remove a single point of compromise or failure. A central authority is removed and replaced by an eco-system of different actors that enact different roles previously performed by that central authority. While DLT is often considered as a method of de facto decentralization (always subject to debate [11]), what appears decentralized at the technical level may not be at all when considering other aspects such as jurisdiction, business agreements, etc. Considering Table I, Sovrin for example, has prioritised the design of an ecosystem of actors and roles for entities that divide how power is shared. This ecosystem, which uses a permissioned distributed ledger as a root of trust that underpins a *web of trust*, provides user choice about which actors take on various roles to safeguard a user identity. Other schemes such as uPort rely on public DLTs for identity claim storage, which also forms the basis of arguments of its decentralization, but central authorities are still relied upon to create identity claims. Thus it appears important to ask which schemes are the most decentralised in practice? It seems unrealistic to expect that digital identity can be delivered without some elements of centralisation.

### B. Evaluate deployability in light of PKI challenges

PKI is a security infrastructure designed to add trust to asymmetric cryptography. It is often informally assumed as a solution-in-waiting for accommodating human identity. However, despite its continuous availability over the previous decades, it is still widely recognised as a complex infrastructure that is used infrequently for human identity. Indeed, back in 1996 Don Davis [8] cited 5 defects in PKI which led to his conclusion that "*these defects make public key cryptography more suitable for server-to-server security than for desktop applications*".

PKI faces challenges due to a number of enduring operational and deployment concerns across the full lifecycle of certification, e.g. certificate issuance and revocation (especially across multiple organisations) [4]. In addition, many industries

do not fit the hierarchical mould imposed by PKI; this has resulted in a large number of independent, often competing, certification authorities, which do not sign each other's keys. Also, unlike online services, which heavily rely on a PKI for secure communication, human identities are not necessarily unique. There might be a unique human being, but that has multiple different and overlapping identities (sets of attributes in different contexts). We are the combination of these distinct identities and we want to be able to alternate between them [14]. In a digital world, this implies being able to derive several sub-identities from a core one.

In prior work, we proposed that *decentralized trusted digital identity* was a mode of DLT-based identity that tends towards a single certification authority that certifies claims that it is hoped others will trust. While *self-sovereign identity* involves users adopting greater operational responsibility for identities including collecting identity claims from an eco-system of identity providers. The latter alleviates reliance on an overarching certification authority but says little about the resulting need to organise federations of certification authorities and their ability to interoperate in terms of policy and technology e.g. mechanisms to validate cross-institution identity claims made about a user. Therefore, one pressing avenue to investigate is how deployability assumptions underpinning new DLT-based identity technologies are born out in practice, and whether these assumptions are better or worse than those made by PKI. Given that asymmetric cryptography underpins both DLT and PKI it remains to be seen how a DLT-based identity application can optimally replace, integrate with, or disregard, principles from PKI architectures.

### C. *Support secure delegation of credentials*

The vision of user-centric identity calls for a future where user control [2] is a defining feature of an identity scheme. One limit of that vision is when users make use of that control, to give it away temporarily, or even over the long-term, to another entity. In short, to delegate that control to another entity. It can already be seen in the cryptocurrency market that reliance upon centralized exchanges can result in compromise of the underlying resources with no recourse possible for the user [7]. Of course, this is simply a newly occurring scenario of central authorities being themselves a vulnerability and mirrors a long-standing threat to centralized ecosystems. Those who occupy a position of trust are able to compromise privacy of their users (via data breach), and could even enable an attacker to masquerade as a user and allow them to leverage associated privileges.

The vision of end-user managed asymmetric cryptography as a dominant means to establish and authenticate end-user identities makes pressing the need for new ways to initiate and revoke delegation capabilities, in ways that address the potential for abuse, and the privacy of the user. Barka and Sandhu [1] describe a framework that captures properties of delegation in the context of operating system access control, in a way that offers transferability to the context of identity. But in general there is an enduring need for two types of delegation: to other *individuals,* and to an *organization*. Related to the former, prior work has documented how access credentials are regularly shared amongst individuals in romantic relationships, and amongst carers that work in care communities for older people [9]. For the latter, technical architectures for digital identity has long held roles for identity providers and identity brokers and given the increased importance of effective key management by end-users, it is likely that the need for *user management* (in the IAM sense) will still be prevalent but take a new form.

It is an open question how DLT can record, or give transparency to delegation of identity credentials and how automated policies can enforce the delegation intentions of a user, but also safeguard the private nature that some of these decisions can exhibit. We should not design digital identity schemes that are even more rigid in their conceptions of one-identity-one-person than systems we have already.

### D. *Gather new requirements for the user experience*

While the collective understanding of user experience in the domain of information privacy and security has improved over the past two decades, very little of that understanding relates to new approaches to support end-user key management, and the apparent mainstream acceptance of cryptocurrencies has not made this issue disappear. Indeed, the words of Davis [8] from 1996 still find relevance as back then he highlighted that users would be unlikely to adopt the required behaviours to successfully manage private keys, and proposed that this task was one of the 5 *compliance defects* inherent to PKI. Losing a private key in the context of digital identity could constitute a major vulnerability, for which throwing a particular pseudonym away and bootstrapping an identity again may not be an acceptable response.

Of course, key management is only one of the challenges. Much work has also considered ways that users can better consent to, and be informed about, sharing their identity data with third parties. The technical nature of DLT presents new challenges to educate users about the persistence of personal data, that go beyond challenges that have so far been experienced in centralized identity infrastructures. Consider, for example, the communication required to explain personal data storage strategies adopted by DLT-based identity schemes: e.g. uPort: credentials stored on Ethereum and represented as plain text data, cryptographic hashes, or cipher-text; Sovrin: no data on the ledger, only a visible ledger that maps identifiers to public keys. How can we best communicate such design decisions and their implications to a user to obtain meaningful consent?

Thinking about usability is one way to consider users in system design, but that technique yields little about whether that system design solves an important problem for users, and thus will be widely used. Recall that Cameron's identity meta-system [3] placed user interface design as a crucial determinant of success, but this did not influence the success of that particular technology. This serves to say that user experience goes deeper than the user interface and that solving important technical problems does not necessarily translate to creating technologies that users want to use. This calls for research to understand end-user experiences of contemporary digital identity mechanisms and to extract design requirements for a new generation of DLT-based approaches.

### E. *Evaluate exposure to public permissionless DLTs*

Public permissionless DLTs are attractive for decentralized application designers since the incentives set by cryptocurrency-based DLTs (typically) make them a stable platform to record small amounts of data over an extended period of time (i.e. the

Bitcoin ledger started 9 years ago). However, one common criticism of these technologies is that the most mainstream instances rely upon proof of work (PoW) to achieve ledger consensus at the waste of large quantities of electricity. In the case of Bitcoin, PoW involves finding a nonce, such that, when hashed with other data on the blockchain, produces a sufficiently small number. While this serves an important function so that peers can immediately recognize that the genuine blockchain is the longest one due to the work involved in building it, it was shown that in 2014 the entire Bitcoin mining operation was on par with Ireland for electricity consumption [16]. Moreover, these schemes often result in an arms race for mining hardware, which effectively leads to centralization of mining capability [11]. While this is evidently a sustainability concern with DLT more generally, the future evolution of digital identity applications in this domain are entangled with the challenge of DLT sustainability since many DLT-based identity schemes rely on these publicly available resources.

DLT-based identity schemes rely upon public permissionless DLTs to differing degrees. For example, uPort relies upon Ethereum smart contracts, while Sovrin has no reliance on public permissionless technologies at all. If the public permissionless reliant schemes see widespread uptake this inevitably increases transaction volume on those technologies, perhaps hindering a transition of those technologies to a more sustainable consensus methods. This transaction load could be made heavier by the fact that in most approaches to DLT-based identity remove the one-to-one mapping between an individual and an identity, resulting in one individual having many identities (or, pseudonyms) to transact with, potentially resulting in more data/credential storage needs. Thus it is important for research to quantify and evaluate the transaction load created on public permissionless DLTs over time and consider balancing an exposure to that resource with other techniques that enable decentralized recordkeeping such as permissioned distributed ledgers or other off-chain storage solutions.

Another consequence of the current design of public and permissionless DLTs is that a fee is taken from each transaction to support the running costs of the transaction miners. Who would bear the transaction costs in the case of a DLT-based identity scheme? Currently, the ability to enrol and interact with online services is largely free for end-users (unless we consider price of acquiring e.g. government credentials). Which transactions should be stored in a DLT (and therefore incur transaction cost) and which should not? If relying on public and permissionless resources the challenge to find the right financial incentives to encourage uptake will be an important task of DLT-based identity scheme designers.

## V. Final Remarks

In this position paper, we considered the application of DLT to digital identity and proposed areas of research that are particularly pressing to gauge future potential. Our focus has been on describing research areas that – in our view – hold a key to better understand how DLT can be put to work in the application area of digital identity; at the expense of other important engineering challenges intrinsic to DLTs themselves (e.g., scalability or reliability). But that is not to say that we under-value the importance and difficulty of those challenges. Future research can considerto target sub-domains of the areas that we have proposed, and even to extend our proposals. Only time – and more research focused on specific use cases – will tell whether DLT can form a useful component of known frameworks of digital identity, or will lead to new frameworks entirely.

## References


1. Ezedin Barka and Ravi Sandhu. 2000. Framework for role-based delegation models. In *Proceedings of the 16th Annual Computer Security Applications Conference* (ACSAC '00). Retrieved from http://portal.acm.org/citation.cfm?id=784591.784743
2. Abhilasha Bhargav-Spantzely, Jan Camenisch, Thomas Gross, and Dieter Sommer. 2006. User centricity: a taxonomy and open issues. In *Proceedings of the second ACM workshop on Digital identity management - DIM '06*, 1. http://doi.org/10.1145/1179529.1179531
3. Kim Cameron and Michael B. Jones. 2007. Design Rationale behind the Identity Metasystem Architecture. In *ISSE/SECURE 2007 Securing Electronic Business Processes*. Vieweg, Wiesbaden, 117–129. http://doi.org/10.1007/978-3-8348-9418-2_13
4. Elias G. Carayannis and Eric Turner. 2006. Innovation diffusion and technology acceptance: The case of PKI technology. *Technovation* 26, 7: 847–855. http://doi.org/10.1016/J.TECHNOVATION.2005.06.013
5. Melissa Chase and Sarah Meiklejohn. 2016. Transparency Overlays and Applications. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (CCS '16), 168–179. http://doi.org/10.1145/2976749.2978404
6. David L. Chaum and David L. 1981. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM* 24, 2: 84–90. http://doi.org/10.1145/358549.358563
7. Coinbase. 2017. Centralised Exchanges Are Terrible At Holding Your Money: A Timeline of Catastrophes. Retrieved January 12, 2018 from https://blog.localethereum.com/centralised-exchanges-are-terrible-at-holding-your-money/
8. Don Davis. 1996. Compliance defects in public-key cryptography. In *Proceedings of the 6th conference on USENIX Security Symposium, Focusing on Applications of Cryptography - Volume 6*, 17. Retrieved from http://dl.acm.org/citation.cfm?id=1267569.1267586
9. Paul Dunphy, Andrew Monk, John Vines, Mark Blythe, and Patrick Olivier. 2013. Designing for Spontaneous and Secure Delegation in Digital Payments. *Interacting with Computers* 10.1093/iw. http://doi.org/10.1093/iwc/iwt038
10. Paul Dunphy and Fabien A.P. Petitcolas. 2018. *A First Look at Identity Management Schemes on the Blockchain*. Retrieved from arxiv:1801.03294 [cs.CR]
11. Adem Efe Gencer, Soumya Basu, Ittay Eyal, Robbert van Renesse, and Emin Gün Sirer. 2018. Decentralization in Bitcoin and Ethereum Networks. Retrieved January 16, 2018 from http://arxiv.org/abs/1801.03998
12. Global Initiative Against Transnational Organised Crime. 2017. The Trouble with Transparency: Increased identity fraud in Sweden's digital age. Retrieved from http://globalinitiative.net/transparency_sweden/
13. Garrick Hileman and Michel Rauchs. 2017. *2017 Global Blockchain Benchmarking Study*. Retrieved from https://ssrn.com/abstract=3040224
14. Ignacio Mas and David Porteous. 2015. *Minding the Identity Gaps*. http://doi.org/http://dx.doi.org/10.2139/ssrn.2189989
15. Satoshi Nakamoto. 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System*.
16. Karl J O'Dwyer and David Malone. 2014. Bitcoin mining and its energy footprint.
17. Al Pascual, Kyle Marchini, and Sarah Miller. 2017. *Identity Fraud: Securing the Connected Life*.
18. Thomson Reuters. 2016. Thomson Reuters 2016 Know Your Customer Surveys Reveal Escalating Costs and Complexity. Retrieved October 5, 2016 from http://thomsonreuters.com/en/press-releases/2016/may/thomson-reuters-2016-know-your-customer-surveys.html
19. Certificate Transparency. Retrieved January 1, 2018 from https://www.certificate-transparency.org/