The University of Texas at Austin
Center for Identity

How Much Identity Management with Blockchain Would Have Saved Us? A Longitudinal Study of Identity Theft

Razieh Nokhbeh Zaeem
K. Suzanne Barber

UTCID Report #20-14

July 2020

How Much Identity Management with Blockchain Would Have Saved Us? A Longitudinal Study of Identity Theft

Abstract

The use of blockchain for identity management (IdM) has been on the rise in the past decade. We present the first work to study the actual, large-scale impact of using blockchain for identity management, particularly how it can protect Personally Identifiable Information (PII) to curb identity theft and fraud. Our insight is that if blockchain-based IdM protects PII, it can reduce the number of theft and fraud cases that take advantage of such PII. At the Center for Identity at the University of Texas at Austin, we have modeled about 6,000 cases of identity theft, and PII exploited in them. We utilize this model to investigate how three real-world blockchain-based IdM solutions (Civic, ShoCard, and Authenteq) could have reduced the identity theft loss over the past 20 years if they had been universally used. We identify which PII protected by blockchain is more critical. We also suggest new PII to include in blockchain-based IdM. Our work paves the way for the design of more effective blockchain-based IdM or any other new line of IdM for that matter.

Key Takeaways

- The Center for Identity present's the first work to study the actual, large-scale impact of using blockchain for identity management, particularly how it can protect Personally Identifiable Information (PII) to curb identity theft and fraud.
- UT CID has modeled approximately 6,000 cases of identity theft, along with the PII exploited in them.
- This model was then used to investigate how three real-world blockchain-based IdM solutions (Civic, ShoCard, and Authenteq) could have reduced the identity theft loss over the past 20 years.
- The research finds that that some PII are more important to protect: complete protection of Social Security Card, for example, would have eliminated about 10% of all the incidents we recorded and modeled.
- This report also recommends new PII to add to blockchain-based IdM solutions: most notably Financial Information, e.g., Debit/Credit Card, Bank Account, and Taxpayer Information.

How Much Identity Management with Blockchain Would Have Saved Us? A Longitudinal Study of Identity Theft

Razieh Nokhbeh Zaeem[1][0000−0002−0415−5814] and K. Suzanne Barber[1]

The Center for Identity, The University of Texas at Austin, Austin, TX 78712, USA
{razieh,sbarber}@identity.utexas.edu
https://identity.utexas.edu

Abstract. The use of blockchain for identity management (IdM) has been on the rise in the past decade. We present the first work to study the actual, large-scale impact of using blockchain for identity management, particularly how it can protect Personally Identifiable Information (PII) to curb identity theft and fraud. Our insight is that if blockchain-based IdM protects PII, it can reduce the number of theft and fraud cases that take advantage of such PII. At the Center for Identity at the University of Texas at Austin, we have modeled about 6,000 cases of identity theft, and PII exploited in them. We utilize this model to investigate how three real-world blockchain-based IdM solutions (Civic, ShoCard, and Authenteq) could have reduced the identity theft loss over the past 20 years if they had been universally used. We identify which PII protected by blockchain is more critical. We also suggest new PII to include in blockchain-based IdM. Our work paves the way for the design of more effective blockchain-based IdM or any other new line of IdM for that matter.

Keywords: Identity Management · Blockchain · Identity Theft · Self-Sovereign Identity.

1 Introduction

The blockchain technology has found a variety of applications beyond cryptocurrency, from insurance to the Internet of Things. One of these applications is Identity Management (IdM), the framework that identifies, authenticates and authorizes users in order to control access to resources. Many concrete blockchain-based IdM solutions exist today, ranging from successful startups to open source projects and foundations.

One of the most prominent goals of IdM, when utilized to manage Personally Identifiable Information (PII) (such as email addresses, passwords, fingerprints, and driver's licenses), is to protect PII from identity theft and fraud. Identity theft and fraud is the fraudulent acquisition and use of such PII. According to the latest statistics from the U.S. Department of Justice published in 2019 [8],

10% of all U.S. residents reported that they had been victims of identity theft in 2016 with a total loss of $17.5 billion.

Blockchain-based IdM solutions offer user control, decentralization, immutability, transparency, security, and privacy for the PII they manage–added properties that can prevent identity theft and fraud. The actual effect of these solutions on the landscape of identity theft and fraud, however, has not received the attention it deserves.

We present the first work to examine the large-scale effect of this new type of IdM through the lens of a longitudinal study of close to 6,000 cases of identity theft and fraud, which took place over the past 20 years. We estimate the potential of blockchain-based IdM in preventing identity theft and fraud, based on its ability to protect PII leveraged in identity management. Our insight is that if blockchain-based IdM secures and protects some types of PII, it can eliminate (or substantially reduce) the number of theft and fraud cases that take advantage of those types of PII. Consequently, it can prevent millions of dollars of loss that harm identity theft victims every year.

We make the following contributions:

1. We introduce the idea of extrapolating from previous identity theft and fraud cases to study the future effect of blockchain-based IdM.
2. We take advantage of our longitudinal study of the past 20 years of news-reporting on identity theft and fraud—modeled by a team of modelers at the University of Texas at Austin Center for Identity in over six years. We investigate the frequency, monetary loss, and other properties of identity theft cases involving PII that could be protected with blockchain-based IdM.

We discover that while blockchain-based IdM cannot eliminate identity theft and fraud altogether, its effective protection of PII such as Social Security Card can potentially end millions of dollars of identity theft loss. The protection blockchain-based IdM solutions extend to *some* PII is more valuable than others. Reducing the chance of exposure of these PII (Social Security Card, Healthcare ID and Driver's License) would have saved identity theft victims from the highest amount of loss in the past 20 years.

Based on our study of the identity theft cases, we recommend new PII to include in blockchain-based IdM to further mitigate the effect of identity theft. Some of the most prominent PII we suggest include: Financial Information such as Banking, Credit/Debit Card, and Taxpayer Information as well as Employer Information and ID.

Our work sheds light on actual, large-scale potential of new IdM solutions, with a focus on blockchain-based IdM. Our results enlighten developers of new IdM and their users and scientifically direct these efforts to maximize their impact. Furthermore, once new blockchain-based IdM is widely used, we can study its actual effect on the landscape of identity theft and fraud through the same methodology of examining reported identity theft cases.

2 Related Work: Identity Management with Blockchain

In this section we explain fundamentals of blockchain and blockchain-based IdM, review commercially available examples of such IdM solutions, and cover related work on the actual large-scale benefits of migrating to IdM with blockchain.

2.1 Blockchain

The term blockchain is rooted in the seminal Bitcoin white-paper [12] that introduced a novel crypto-currency (i.e., electronic cash) technology. This technology allows online transactions to take place without the need to go through a trusted financial third party. Digital signatures and a peer-to-peer network form the backbone of the Bitcoin technology. The two parties of the transaction communicate through digital signatures (i.e., public and private keys). The peer-to-peer network timestamps transactions by hashing them into a chain of blocks, forming a record of transactions. The longest chain of blocks serves as a tamper-proof ledger of all witnessed transactions. This ledger (also known as blockchain) cannot be altered without the consensus of the network majority.

2.2 Blockchain-Based Identity Management

The blockchain technology has been used for a whole host of applications [10], including identity management (IdM). IdM is the framework that identifies, authenticates, and authorizes users to access resources. Blockchain-based IdM solutions [18] adopt blockchain for identity management.

In an IdM, when the user needs to make a *claim* (i.e., assert something about one's identity, like the citizenship of a country), he/she provides an *identity proof* (i.e., some form of document that provides evidence for the claim, like a passport). Identity proofs always contain PII, i.e., any information that could be used to identify an individual. Such identity proofs should be *attested* (i.e., validated) by the relevant identity authority (e.g., the agency that issued the passport).

Blockchain-based identity solutions encrypt a user's identity, hash it, and add its attestations to the blockchain ledger. These attestations are later used in order to prove the user's identity. Two categories of blockchain-based IdM solutions exist [7]:

1. Decentralized Identity (e.g., **ShoCard**, **Authenteq**, BitID, ID.me, and IDchainZ): This identity solution is similar to conventional identity management solutions where credentials from a trusted service are used. The only difference arises is the storage of validated attestations on a distributed ledger for later validation by a third party.
2. Self-sovereign identity (e.g., **Civic**, Sovrin, uPort, and Onename.io): The user owns and controls his/her identity without heavily relying on central authorities. In essence, self-sovereign identity is very similar to how non-digital identity documents work today. Every user keeps their own identity

documents in their device. The user creates a public/private key pair and contacts identity authorities to associate and attest his/her public key with an identity proof. When the user makes a claim, the user provides his/her attested public key. The verifier accepts only the claims signed with the user's private key.

Blockchain-based IdM improves identity management in several ways. Digital signatures, one of the major components of the blockchain technology, provide authenticity of the identity proof and attestation. The peer-to-peer network, the other major component of blockchain, eliminates the need for a central repository of users' identity. Hence, blockchain can make IdM solutions decentralized, tamper-resistant, and enhance security and privacy.

2.3 Examples of Commercially Available IdM Solutions

Many foundations, companies, startups, and open source projects have utilized [9] blockchain for identity management. In this work, we have studied different blockchain-based IdM services offered by multiple companies like Authenteq [2], ShoCard [3], and Civic [1]. In all these services, the user first logs into the web/mobile app, and provides email address and phone number to create an account. The user then selects which government-issued identity documents to use for identity verification, e.g., Passport, National Identity Card, Driver's License, or Social Security Card as shown in Table 1. The user scans one of these identity verification documents and the app verifies the document against a third party. After the check, the user identity is confirmed, attested on the blockchain and can be reused.

Table 1. Identity document options in three blockchain identity management solutions.

Civic	ShoCard	Authenteq
Passport	Social Security Card (SSC)	Any Government ID
National Identity Card (NID)	Green Card (GC)	
Driver's License	Health Card/Photo ID	

2.4 Evaluation of Blockchain-Based IdM Solutions

There are very few studies evaluating blockchain-based IdM solutions. For example, Baars [4] studied ten blockchain-based identity management systems to identify their properties and compare them together. These ten solutions were Onename.io, Qiy, iDIN, eHerkenning, IRMA, PKIoverheid, Jumio, Tradle, Idensys, and uPort. Dunphy and Petitcolas [7] used the laws of identity framework

(including user control and consent, minimal disclosure, justifiable parties, directed identity, design for a pluralism of operators, human integration, and consistent experience across contexts) to evaluate three blockchain-based IdM solutions (uPort, ShoCard, and Sovrin).

Our previous work [14] investigated different PII options given to users for authentication on current blockchain-based IdM solutions. Based on our Identity Ecosystem model [15, 13, 5, 6, 11], we evaluated these options and their risk and liability of exposure. Powered by real world data of about 6,000 identity theft and fraud stories from ITAP (Section 3), our model recommended some authentication choices and discouraged others.

None of these studies, however, have looked at the potential large-scale effect of this new line of IdM solutions. In this work, we investigate how the added security and privacy of blockchain-based IdM can thwart identity theft. Clearly, using blockchain would not eliminate all identity theft nor it would fully protect identity documents. There still exists a chance that the device containing the identity proof falls into the wrong hands and one poses as the identity owner. With that said, we can still estimate the potential of blockchain-based IdM in reducing the effect of identity theft since it dramatically reduces the chance of compromise for the identity proof/PII. Assuming that blockchain-based IdM protects the identity proof/PII, we approximate how much these IdM solutions would have saved identity theft victims by eliminating identity theft and fraud cases that happened because a particular identity proof/PII was stolen/misused.

3 Identity Threat and Assessment Prediction (ITAP)

In order to estimate the potential of blockchain-based IdM in reducing financial loss of identity theft, we need a database of identity theft and fraud cases that records details of such cases and spans a long period of time and a diverse set of victims. Our Identity Threat and Assessment Prediction (ITAP) project [17, 16] is such a longitudinal study of identity theft and fraud cases.

ITAP is a structured database of news stories that gathers and models data about incidents of identity theft, fraud, and abuse. Through these news stories, we seek to determine the methods and resources used to carry out these crimes, the vulnerabilities exploited, and the consequences of these incidents. The ITAP model is a large and continually growing, structured repository of such information. ITAP spans national and international identity theft cases over the past 20 years (2000-2020) and currently includes 5,906 unique identity theft incidents.

In order to populate ITAP, using a wide variety of online sources, we gather the data from news stories that report on incidents involving the exposure, theft, or fraudulent use of PII such as names, social security numbers, and credit card numbers. We find candidate news stories in two ways: via an RSS feed set up through Feedly[1] and via several Google Alerts[2] based on relevant key phrases.

[1] https://feedly.com
[2] https://www.google.com/alerts

We regularly monitor the stories thus gathered and store those that we deem appropriate for ITAP.

A team of modelers record salient information about these incidents. The recorded information about identity thieves includes:

1. Performers: the performers of the identity theft incident, including thief, frustrater, abuser, or non-malicious actor.
2. Inputs: PII that the performers initially obtained and used.
3. Outputs: PII that the performers created, acquired, or exposed based on the inputs.
4. Resources: The types of tools, devices, applications, and other instruments used by the performers in carrying out the incident.
5. Steps: The steps that the performers took.
6. Criminal Activities: The relationship between the performers and the victims, e.g., insider, outsider, or both.

The information collected about victims includes:

1. Age Group of Victims.
2. Gender of Victims.
3. Citizenship of Victims.
4. Education Level of Victims.
5. Annual Income of Victims.
6. Profession of Victims.
7. Organization Affected.
8. Sector of Society or Infrastructure Involved.
9. Counter Measures Taken by the Victim Organization.

Finally, the information about the incident itself includes:

1. Location of the Event, or the "Internet".
2. Date When Event Occurred.
3. Date of Article or Announcement.
4. Type of Loss Incurred, e.g., emotional, reputation, or financial.
5. Financial Loss Amount (converted to US Dollars).
6. Reputation Loss.
7. Emotional Distress.

4 Experimental Results

In this work, we seek to answer the following main research question: What is the potential of the blockchain-based identity management solutions (namely Civic, ShoCard, and Authenteq) in thwarting identity theft? For example, how much financial damage would have been avoided if all the ITAP identity theft and fraud cases that have a Social Security Number as input where eliminated, because the owner protected his/her Social Security Number with blockchain-based identity management? In order to answer this question, we report statistics from ITAP to answer the following research questions (RQ):

1. How many cases have at least one PII input that could be protected with blockchain-based IdM (Table 1)?
2. What is the average financial cost of those cases?
3. What are other PII that could be added to blockchain-based IdM to prevent
   (a) high-loss identity theft incidents?
   (b) frequent identity theft incidents?

We retrieved the list of PII involved in ITAP identity theft and fraud cases (i.e., all PII that thieves and fraudsters used as inputs to carry out the incidents or expose/generate other PII as well as all PII that they exposed or fraudulently generated as outputs) and manually scanned them for relevant PII according to Table 1. Table 2 lists these PII and shows which IdM solution could possibly protect them.

4.1 RQ1: How many cases have at least one input PII that could be protected with blockchain-based IdM?

There are a total of 5,906 identity theft and fraud cases in ITAP, and Table 3 shows how many of those cases have at least one input PII from a given category of PII. (Categories of PII are shown in Tables 1 and 2.) Note that Green-Card-related PII (e.g., Visa Details) happen to appear only as *output* in the current set of ITAP cases, and therefore have zero cases in which they are PII *inputs*. Also, there are only two cases with National ID as their input PII, none of which report their financial loss. As a result, we cannot calculate the average loss for the Green Card and National ID categories of PII. This might be due to the fact that ITAP is predominantly U.S.-oriented, where Visa and National ID contribute to a very small number of identity theft cases.

It is clear from the number of cases reported in Table 3 compared to the total number of 5,906 cases in ITAP that blockchain-based IdM solutions do not eliminate all identity theft and fraud, even if universally applied to protect the PII these solutions do cover. These IdM solutions, however, would save the identity theft and fraud victims from considerable amount of financial loss.

4.2 RQ2: What is the average financial cost of preventable cases?

Most of the PII categories of Table 3 pertain to an average loss amount of over $1M (not averaged per victim but per incident[3]). For example, if all PII of the Passport category (including Passport Information, Number, Expiration Date, and Country of Issue) where protected with blockchain-based IdM, 16 identity theft cases of ITAP with an average loss value of $1,252,464 would have been avoided because at least one of their input PII would not be compromised.

The Social Security Card PII category (including Social Security Number, Suffix, and Social Security Card) is the input to 567 cases with an average

[3] We did not calculate the average loss per victim because the number of victims is usually not reported in the identity theft and fraud news story.

Table 2. PII that could be protected with the blockchain-based IdM solutions Civic, ShoCard, and Authenteq, as found in the ITAP repository.

Authenteq	ShoCard	Insurance Policy Information Health Plan Member Number Health Insurance Policy Number Individual Healthcare Plan Health Insurance Policy Information Healthcare Provider's Name Medical ID Number Healthcare Provider Information Health Insurance ID Card Information Health Plan Group Number Health Insurance Company Name	Private Health ID
		Medicare Provider ID Number Medicaid ID Number Medicare ID Number Medicaid Provider ID Number National Health Index Number (New Zealand)	Gov. Health ID
		Social Security Number (SSN) SSN Suffix (Last Four Digits) Social Security Number - Invalid	SSC
		Visa Details	GC
	Civic	Passport Information Passport Number Passport Expiration Date Passport Country of Issue	Passport
		National Identity Number	NID
		Driver's License Number Driver's License Information Stolen Driver's License Information Fake Driver's License Information Driver's License Photo Driver's License Expiration Date	Driver's License
		Government Issued ID Number State Identification Number ID Card Number Universal ID Number ID Card Information Aadhaar Unique Identification Number (India) Personal Record Identifier (PRI) (Canada) DNI (Documento Nacional de Identidad - Spain) Information	Other Gov. ID

financial loss of over $27M. To provide some context, Social Security Number is one of the PII involved in the highest loss value cases: the average loss value of all the ITAP cases is just over $11M, while Social Security Card and Number cases have an average loss of over $27M.

4.3 RQ3: What are other PII to add to blockchain-based IdM?

Top 5% PII/identity proof documents in terms of feeding input to high-loss identity theft incidents are as follows:

- Financial Information, including Credit Card Information, and Taxpayer Information such as Electronic Filing Identification Number (EFIN)
- Social Security Number
- Employer Information and ID
- Medicare ID Number

Social Security Number and Medicare are already included in many blockchain-based IdM solutions. We recommend further addition of Credit Card Information and Taxpayer Information, as well as Employer Identification to avert high-loss identity theft and fraud.

Top 5% PII/identity proof documents frequently involved as input to incidents are as follows:

- Social Security Number
- Financial Information, e.g., Credit/Debit Card, Bank Account, and Taxpayer Information
- Driver's License Information
- Patient Medical Record

These frequently abused PII have a considerable overlap with PII involved in high-loss incidents. We reiterate that one category of PII/identity proof that we strongly recommend adding to blockchain-based IdM is Financial Information, e.g., Debit/Credit Card, Bank Account, and Taxpayer Information. In fact, the investigation of the websites of Authenteq, Civic, and ShoCard shows they are planning to include this information in their IdM solutions.

Table 3. The number of ITAP cases that have a PII protected with the blockchain-based IdM solutions Civic, ShoCard, and Authenteq, as *input*.

IdM		Authenteq						
	ShoCard				Civic			
PII Cat.	Private Health ID	Gov. Health ID	Social Security Card	Green Card	Passport	National ID	Driver's License	Other Gov. ID
# Cases	28	19	567	0	16	2	86	42
Avg. Loss	$1,661,308	$4,052,047	$27,465,086	N/A	$1,252,464	N/A	$1,751,281	$492,459

5 Conclusion and Future Work

We presented the first work in the literature that investigated the real-world, large-scale impact of blockchain-based IdM on identity theft and fraud. We utilized a model of about 6,000 reported identity theft and fraud cases that occurred in the past 20 years (from 2000 to 2020), and estimated how much blockchain-based IdM solutions (Civic, ShoCard, and Authenteq) could have saved identity theft victims, if they had been used. We found that some PII are more important to protect: complete protection of Social Security Card, for example, would have eliminated about 10% of all the incidents we recorded and modeled, with an average monetary loss of over $27M per incident. We also recommended new PII to add to blockchain-based IdM solutions: most notably Financial Information, e.g., Debit/Credit Card, Bank Account, and Taxpayer Information. Our results are useful to developers of new this new line of IdM and their users. In the future, we can retrospectively evaluate the effect of the universal use of blockchain-based IdM through the same methodology, once it is widely adopted.

6 Acknowledgments

This work was in part funded by the Center for Identity's Strategic Partners. The complete list of Partners can be found at https://identity.utexas.edu/strategic-partners.

References

1. Civic decentralized reusable kyc services - blockchain-powered. https://www.civic.com/solutions/kyc-services/, (Accessed on 04/11/2019)
2. Identity verification & KYC, authenteq. https://authenteq.com/, (Accessed on 04/11/2019)
3. Shocard identity management use cases — shocard. https://shocard.com/identity-management-use-cases/, (Accessed on 04/11/2019)
4. Baars, D.: Towards self-sovereign identity using blockchain technology. Master's thesis, University of Twente (2016)
5. Chang, K.C., Zaeem, R.N., Barber, K.S.: Enhancing and evaluating identity privacy and authentication strength by utilizing the identity ecosystem. In: Proceedings of the 2018 Workshop on Privacy in the Electronic Society. pp. 114–120 (2018)
6. Chang, K.C., Zaeem, R.N., Barber, K.S.: Internet of things: Securing the identity by analyzing ecosystem models of devices and organizations. In: 2018 AAAI Spring Symposium Series. pp. 111–116 (2018)
7. Dunphy, P., Petitcolas, F.A.: A first look at identity management schemes on the blockchain. IEEE Security & Privacy **16**(4), 20–29 (2018)
8. Erika Harrell, Ph.D., B.o.J.S.: Victims of identity theft (2016, revised in 2019)
9. Jacobovitz, O.: Blockchain for identity management. The Lynne and William Frankel Center for Computer Science Department of Computer Science. Ben-Gurion University, Beer Sheva (2016)
10. Lakhani, K.R., Iansiti, M.: The truth about blockchain. Harvard Business Review **95**, 118–127 (2017)

11. Liau, D., Zaeem, R.N., Barber, K.S.: Evaluation framework for future privacy protection systems: A dynamic identity ecosystem approach. In: 2019 17th International Conference on Privacy, Security and Trust (PST). pp. 1–3. IEEE (2019)
12. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2008). Tech. rep., Manubot (2019)
13. Rana, R., Zaeem, R.N., Barber, K.S.: Us-centric vs. international personally identifiable information: A comparison using the ut cid identity ecosystem. In: 2018 International Carnahan Conference on Security Technology (ICCST). pp. 1–5. IEEE (2018)
14. Rana, R., Zaeem, R.N., Barber, K.S.: An assessment of blockchain identity solutions: Minimizing risk and liability of authentication. In: 2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI). pp. 26–33. IEEE (2019)
15. Zaeem, R.N., Budalakoti, S., Barber, K.S., Rasheed, M., Bajaj, C.: Predicting and explaining identity risk, exposure and cost using the ecosystem of identity attributes. In: 2016 IEEE International Carnahan Conference on Security Technology (ICCST). pp. 1–8. IEEE (2016)
16. Zaeem, R.N., Manoharan, M., Yang, Y., Barber, K.S.: Modeling and analysis of identity threat behaviors through text mining of identity theft stories. Computers & Security **65**, 50–63 (2017)
17. Zaiss, J., Nokhbeh Zaeem, R., Barber, K.S.: Identity threat assessment and prediction. Journal of Consumer Affairs **53**(1), 58–70 (2019)
18. Zyskind, G., Nathan, O., et al.: Decentralizing privacy: Using blockchain to protect personal data. In: 2015 IEEE Security and Privacy Workshops. pp. 180–184. IEEE (2015)

WWW.IDENTITY.UTEXAS.EDU