An Assessment of Blockchain Identity Solutions: Minimizing Risk and Liability of Authentication

Rima Rana
Center for Identity
The University of Texas at Austin
Austin, Texas, USA
rima.rana@utexas.edu

Razieh Nokhbeh Zaeem
Center for Identity
The University of Texas at Austin
Austin, Texas, USA
razieh@identity.uteaxs.edu

K. Suzanne Barber
Center for Identity
The University of Texas at Austin
Austin, Texas, USA
sbarber@identity.utexas.edu

ABSTRACT

Personally Identifiable Information (PII) is often used to perform authentication and acts as a gateway to personal and organizational information. One weak link in the architecture of identity management services is sufficient to cause exposure and risk identity. Recently, we have witnessed a shift in identity management solutions with the growth of blockchain. Blockchain—the decentralized ledger system—provides a unique answer addressing security and privacy with its embedded immutability. In a blockchain-based identity solution, the user is given the control of his/her identity by storing personal information on his/her device and having the choice of identity verification document used later to create blockchain attestations. Yet, the blockchain technology alone is not enough to produce a better identity solution. The user cannot make informed decisions as to which identity verification document to choose if he/she is not presented with tangible guidelines. In the absence of scientifically created practical guidelines, these solutions and the choices they offer may become overwhelming and even defeat the purpose of providing a more secure identity solution.

We analyze different PII options given to users for authentication on current blockchain-based solutions. Based on our Identity Ecosystem model, we evaluate these options and their risk and liability of exposure. Powered by real world data of about 6,000 identity theft and fraud stories, our model recommends some authentication choices and discourages others. Our work paves the way for a truly effective identity solution based on blockchain by helping users make informed decisions and motivating blockchain identity solution providers to introduce better options to their users.

WI '19, October 14–17, 2019, Thessaloniki, Greece

ACM ISBN 978-1-4503-6934-3/19/10…$15.00
https://doi.org/10.1145/3350546.3352497

CCS CONCEPTS

• **Security and privacy → Privacy protections**; • **Social and professional topics → Privacy policies**; • **Applied computing →** *Digital cash.*

KEYWORDS

Identity, Blockchain, Privacy, Privacy Policy, Authentication, Personally Identifiable Information, Identity Ecosystem

ACM Reference Format:
Rima Rana, Razieh Nokhbeh Zaeem, and K. Suzanne Barber. 2019. An Assessment of Blockchain Identity Solutions: Minimizing Risk and Liability of Authentication. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI '19), October 14–17, 2019, Thessaloniki, Greece.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3350546.3352497

1 INTRODUCTION

According to the Identity Theft Resource Center's report, there was a 126% increase in the number of breached records that contained sensitive Personally Identifiable Information (PII) from 2017 to 2018 [1]. PII is defined as any information used to distinguish or trace the identity of an individual [5]. These unique identity traits associated with an individual are core to our physical and digital lives. PII can be online like an email address or physical like a driver's license. These attributes put together help to differentiate one person from another.

Identity breaches can happen in the most unexpected places. Many victims reported incidents in which PII were compromised through third party vendors [1], such as Identity management (IdM) systems. With such a high rate of identity breaches, it becomes imperative to address the issue of risk involved with PII in Identity management systems.

IdM systems based on the blockchain technology have been on the rise in the past decade. A blockchain is a shared database or decentralized distributed ledger. Bitcoin ledger was the first blockchain but now many more exist across the world. A blockchain consists of chains of "blocks" containing

information about transactions, participants, and identifiers. The features that make blockchain very interesting are immutability, security, and transparency.

In blockchain-based identity solutions, different applications and services are available. A self sovereign identity solution gives the control of one's personal information to the individual in a more secure environment [13]. The user first creates an account with the blockchain identity service and then logs into the account. Here the user provides his/her choice of government attested identity document used for verification which is obfuscated to create blockchain attestations [3, 4, 6]. The IdM solution offers the user multiple choices of documents that are acceptable. The user, however, cannot make an informed decision about which identity verification document to use without proper guidelines. Yet, blockchain-based IdM solutions do not provide such guidelines.

In this paper, we help the user of a blockchain-based IdM in making informed decisions about what identity documents to use on the IdM. We achieve this goal by evaluating the current blockchain-based identity verification solutions with the help of a statistic graphical model of identity—Identity Ecosystem [14] —which covers more than 6000 identity theft stories from across the world [12].

The Identity Ecosystem is a highly sophisticated graphical model of identity attributes and their relationships for people, devices and organizations. Each attribute is modelled as a node and relationship between the attributes as edges. It provides a framework to predict the risk of losing PII and the liability associated with the fraudulent use of respective PII attributes. It can also be used to estimate the monetary value of PII attributes, and determine the connectedness and dependencies between PII. The Identity Ecosystem tool is using Bayesian inference to perform three types of queries: 1) analyzing the risk of exposure, 2) inferring the most likely source of a breach, and 3) calculating the expected cost of attributes.

In order to provide guidelines to users about choosing identity verification documents wisely in the current blockchain identity solutions, we are using the PII nodes' graph properties and relationships and the first query of Identity Ecosystem. We perform experiments and analyze the identity sets required in specific use cases to recommend more secure authentication choices and avoid risky and/or costly PII. This work provides a holistic picture for the user to make informed decisions on blockchain-based IdM systems and inspires the providers of these systems to improve their IdM solutions.

The paper is organized as follows: Section 2 deals with the related work covering blockchain, identity management with blockchain and identity verification solutions based on blockchain. Section 3 provides a brief description of our methodology involving the comprehensive identity framework model, Identity Ecosystem, and our two approaches using Identity Ecosystem. Section 4 deep dives into comprehensive analysis for evaluation. Finally, Sections 5 and 6 conclude our research and provide insights for future works.

2 RELATED WORK

Blockchain And Its Advantages

Blockchain is a new technology supporting distributed ledgers that does not need a central authority to validate transactions; instead these transactions are shared amongst peers [10, 11]. It has a different type of consensus mechanisms in place to achieve its state of transactions in its record, named the ledger [8]. Apart from cryptocurrency, blockchain has found many applications including identity management. As previously suggested in related work [2, 9], the advantages of using blockchain to IdM are as follows:

- Decentralization: Identity stored in a ledger is not controlled by a single authority.
- Immutability: As transactions are appended only and verified by all members, its integrity can be checked by anyone.
- Transparency: Everything on the ledger is visible to everyone.
- Security: As the blockchain is maintained and verified by so many actors, no one can influence the state without getting the majority.

Identity Management With Blockchain

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. It also incorporates the following different ways [9]. This section covers the important schemes/concepts used in these solutions.

- Decentralized Identity: This identity solution is similar to the 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.
- Self-sovereign identity: A person is the one who owns and controls his/her identity without heavily relying on central authorities. It provides a framework to enable exchange of information and propagation of trust between peers.
- Zero knowledge proofs: It is a cryptographic method by which one party (the prover) can prove to another party (the verifier) that they know a value x, without conveying any information apart from the fact that

Table 1: Identity Document options in Blockchain IdM Solutions.

Civic	ShoCard	Authenteq
Passport	Social Security Card	Any government Id
National Identity Card	Green Card	
Driver's License	Health Card/Photo Id	

they know the value x. It is used in blockchain to perform authentication without giving the secret to other party [7].

Blockchain Identity Verification Solutions

Our research focuses on the blockchain-based identity verification solutions. We have studied different blockchain-based IdM services offered by multiple companies like Authenteq [4], ShoCard [6], and Civic [3]. In all these services, the user first logs in to 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, Driving License, or Social Security Number 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. Our research seeks to help the user by providing guidelines to answer the following question: "For blockchain-based identity verification, which is the best set of identity attributes to use to minimize the risk of exposure and loss value while completing the authentication process?" It is important to note that identity attributes are being shared with third parties and also data is stored in the individual's device creating a risk for breach [1]. Presence of unsafe mobile apps and malware installed on a device can trigger a breach in the entire mobile device risking the PII used in the blockchain-based IdM service.

How Is blockchain-based Identity Different Than Current Identity?

Digital Identity is one of the biggest problems on the internet and still there are a lot of security issues surrounding proving a digital identity. However, with the concept of self-sovereign identity being implemented with blockchain, there is a change in the actors involved in controlling identity rather than just the technical system process changing [9]. Now the individual is able to control his/her identity by choosing which PII attributes to use for authentication and also storing the personal information hashed securely on blockchain, available to everyone. This contrasts with the present system where a person hands over the identity document to a third party. The third party controls and stores these documents in a central database and sometimes shares it without the individual's knowledge. Yet, it is not sufficient alone to produce a technologically better identity solution, if the user is not properly trained how to best use this technology.

With the blockchain-based identity, the extent of control over identity increases with which an individual can make more informed decisions about which attributes to use for proving identity. This leads to the bigger question of how can an individual make a better decision in choosing which identity documents to use to prove his/her identity on blockchain-based services. We aim to provide helpful guidelines using Identity Ecosystem with which the risk of exposure and impact of using different PII attributes for identity verification can be compared.

3 METHODOLOGY

In this section, we seek to provide guidelines about choosing identity documents best suited for authentication while minimizing the risk and liability for blockchain-based identity verification solutions. We first explain our previous work, the probabilistic tool Identity Ecosystem, focusing on the graphical model and query used for our research. We then define two approaches using the Identity Ecosystem to apply to the PII attributes from these solutions.

Identity Ecosystem

To understand our approach, first we need to understand the Identity Ecosystem which is developed at the Center for Identity at the University of Texas at Austin (UT CID). It is a graph-based identity model wherein the nodes represent an individual's identity attributes and the edges represents the relationships between these nodes as shown in Figure 1 [14]. It provides a statistical framework for understanding the value, risk and mutual relationships of personally identifiable information attributes. It stores known data about identity attributes in a probabilistic model and performs Bayesian Network-based inference to calculate the posterior effects on each attribute. The Identity Ecosystem is represented as a graph G(V, E) consisting of N attributes A_1, ...,A_N and a set of directed edges as a tuple $e_{i,j} = < i, j >$ where A_i is the originating node and A_j is the target node such that $1 \leq i, j \leq N$ as shown in Figure 1. Each edge $e_{i,j}$ represents a possible path by which A_j can be breached given that A_i is breached.

The Identity Ecosystem is capable of answering questions relevant to the overall risk and liability of any person in terms of managing identity attributes. The query used in our approach is the first one, "When a set of attributes is exposed, how does it affect the risk of other attributes being exposed?" The model can also be utilized to compute changes in probability of breach at other nodes, when data at a node is compromised. Bayesian inference model is utilized to compute this change. Multiple nodes can be selected as

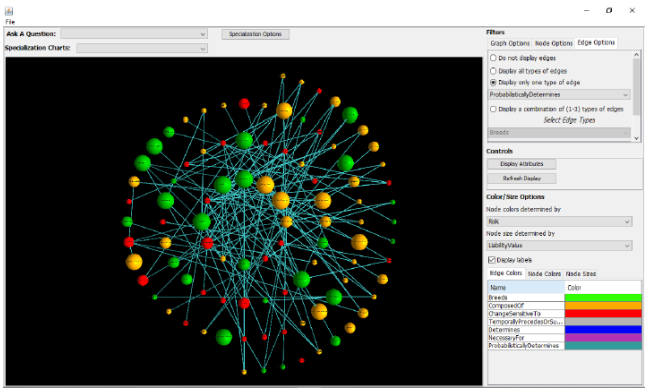

Figure 1: A snapshot of identity attributes graph in Ecosystem.

evidence (i.e., breached PII) and the result shows potential loss after breach for all the affected nodes. For example, if the Social Security Number (SSN) and Social Security Card of an individual are compromised, what set of nodes are most at risk? The answer to this query with potential monetary loss is shown in Figure 2.

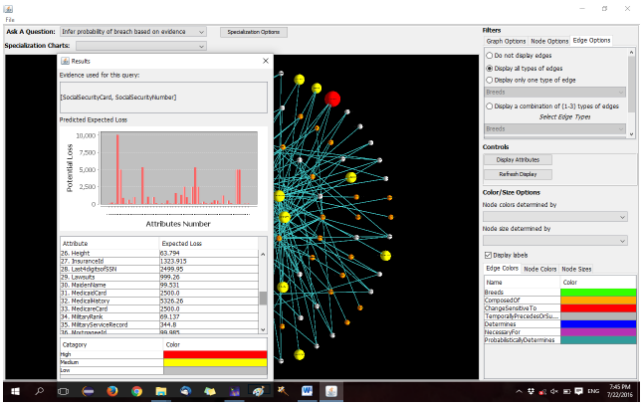

Figure 2: A snapshot of Query "Infer probability of breach based on evidence".

Graph Statistics Approach

Studying three popular blockchain-based identity verification solutions, we observed the use of a limited set of PII (identity documents) for authentication in them [3, 4, 6]. The PII used in these solutions are Email Address and Phone Number for account creation or enrollment and either of these government issued identity cards: Social Security Number, Driver's License Number, Passport Information, and National Identity Number.

We use the Identity Ecosystem to investigate different properties associated with these PII nodes. As we discussed

Table 2: Prior Probability and Intrinsic Loss Value.

PII Node	Prior Probability	Loss (USD)
Social Security Number	0.096598	27,465,086
Driver's License Number	0.008719	2,314,811
Passport Information	0.002565	1,252,465
National Identity Number	0.000342	0
Email Address	0.027526	18,105,024
Phone Number	0.017439	4,405,490

Table 3: Outdegree/Number of Children.

PII Node	Number of Children
Social Security Number	10
Driver's License Number	21
Passport Information	11
National Identity Number	1

in previous sections, each identity attribute is represented as a graph node in Ecosystem. This node has a prior probability, meaning the probability this node is likely to be exposed on its own before the breach evidence set is given. It also has an intrinsic loss value in US Dollars based on the collection of all identity theft and fraud cases researched involving these PII. We tabulate the prior probability and intrinsic loss value for each of the PII nodes frequently used in the blockchain IdM solutions in Table 2. More details on how the data in this table is calculated is provided in Section 4.

The Identity Ecosystem used for this research is modelled as a directed graph with identity attributes as nodes and "probabilistically determines exposure" relationship between them as edges [14]. We tabulate for each of the PII of interest, their outdegree which is the number of children with loss values and indegree which is the number of parents in Tables 3 and 4 respectively. (A child in the Identity Ecosystem is a PII that might be exposed, with a certain probability, if its parent is exposed.) We also show the number of nodes present in a tree rooted at that particular PII node with a depth of 2 in Table 5. The total loss value for different sets of PII attributes used for identity verification is shown in Figure 3. The impact of breach of PII can be determined by its children and its risk of exposure by parents. We analyze all the numbers and determine trends and discuss the results in a later section.

Query Approach

We use the Identity Ecosystem to run the first query "estimating the exposure of breach given a set of identity attributes are exposed". We run this query providing set of evidence from each case mentioned below for the authentication in blockchain-based identity verification solutions. We compare

Table 4: Indegree/Number of Parents.

PII Node	Number of Parents
Social Security Number	27
Driver's License Number	8
Passport Information	4
National Identity Number	2

Table 5: Number of nodes in tree of depth 2.

PII Node	Number of Nodes
Social Security Number	162
Driver's License Number	302
Passport Information	210
National Identity Number	1

the following cases as the email address and phone number are required before choosing any one identity document for verification:

- Email Address + Phone number (Base Case)
- Email Address + Phone number + SSN
- Email Address + Phone number + Passport
- Email Address + Phone number + Driver's License
- Email Address + Phone number + National Identity Card

The goal is to identify which of the above sets, if compromised, introduces a smaller increase in the risk of exposure and the liability value of the individual identity as a whole, including possible future compromises that might stem from that event. With the Bayesian network in Identity Ecosystem, a posterior probability of exposure and loss of monetary value is calculated for all the affected nodes and nodes are classified into high, medium and low categories based on their risk and loss values. The nodes with high risk level after the breach in each case is tabulated and analyzed further in the next section.

4 EVALUATION

In this section, we explain the data source used to populate the Identity Ecosystem and how we extract results from it. Then we extensively discuss the results from both the approaches discussed above (Graph Statistics and Query approaches) and analyze them.

ITAP

The Identity Ecosystem takes its input from the Identity Threat Assessment and Prediction (ITAP). The ITAP is a risk assessment tool which collects case data from sources like law enforcement and the news media. It significantly enhances the understanding of identity theft processes and patterns of threats and vulnerabilities[17]. The ITAP models instances of identity crime and accumulates them to analyze and describe the identity vulnerabilities, the value of identity attributes, and their risk of exposure. The ITAP model describes the business process; comprising of inputs, process steps, outputs, consequences, and victims impacted; by which PII is deliberately stolen, accidentally exposed, and fraudulently used [16]. The ITAP database is structured, elaborate and growing. The repository is covering around 6000 such stories to provide a comprehensive picture of identity theft. The cases analyzed are including the latest cases modeled as of December 2018.

ITAP calculates the Prior Probability of Exposure and Intrinsic Loss Value shown in Table 2. We estimate the risk (i.e., probability) of exposure for any PII i as follows: $risk(i) = \#cases(i)/\#cases$ where $\#cases(i)$ is the number of theft cases in which this PII was exposed to or misused by the criminal and $\#cases$ is the total number of theft cases in ITAP. The dollar value of any PII i is the average dollar value lost in cases where i was misused by identity criminals. Hence, $value(i) = \sum_{c \in cases(i)} v_c/\#cases(i)$ where v_c is the dollar value loss incurred in a given case c[15].

Comparative Analysis

In this section, we demonstrate the results from the two approaches and discuss the insights they provide. At present, there are 627 identity attribute nodes in Identity Ecosystem covering more than 6000 identity theft cases. We first apply the graph statistics approach on the data. We have listed the prior probability and intrinsic loss value for the identity nodes in question for authentication in Table 2. We observe that the prior probability and loss value is highest for Social Security Number. We see Social Security Number has the highest probability of being involved in an identity theft and fraud case, at least 10 times more than any other node considered. From Figure 3, we can observe that the loss value of SSN along with email address and phone number is twice as compared to any other PII set. All these indicate that Social Security Number is at highest risk of exposure and leads to the maximum financial consequences.

We have also listed the number of children for those identity nodes in Table 3. This data shows Driver's License Number has the highest number of children which is double than any other identity node. It will impact the greatest number of nodes if it is exposed. We can also deduce the number of children for National Identity Number is 1 or there are no outgoing edges from it. Hence, it is most safe to use for blockchain-based identity verification in our research question based on this data. To emphasize the trends in vulnerability impact of the selected identity nodes, we list the number of nodes in the tree of depth 2 in Table 5. It further proves our observation where Driver's License Number

Table 6: Count of high and medium risk nodes after running query.

Evidence Node Set	High Risk Node Count	Medium Risk Node Count
Email Address, Phone Number	22	82
Email Address, Phone Number, SSN	22	92
Email Address, Phone Number, Passport	24	88
Email Address, Phone Number, Driver's License	29	96
Email Address, Phone Number, National Identity Card	22	82

Table 7: List of high risk nodes after running query for Email Address and Phone Number.

1. Username	2. Name	3. Expiration Date
4. Address	5. Date of Birth	6. Social Security Number
7. Account Number	8. Credit Card Number	9. Debit Card Information
10. ID Card Information	11. Stolen Driver's License Information	12. Driver's License Number
13. Bank Account Number	14. Monetary Amount	15. CVV Code
16. User Credentials	17. Account Information	18. Physical Address
19. Check Information	20. Patient Medical Record	21. Routing Number
22. Insurance Policy Information		

Table 8: List of high risk nodes after running query for Email Address, Phone Number and SSN.

1. Password	2. Username	3. Bank Account Information
4. Name	5. Address	6. Date of Birth
7. Debit Card Information	8. ID Card Information	9. Stolen Driver's License Information
10. Monetary Amount	11. W-2 Form Information	12. CVV Code
13. User Credentials	14. Employee Record	15. Fake ID Card Information
16. Account Information	17. Personal Identification Number (PIN)	18. Biographic Data
19. Check Information	20. Birth Certificate Information	21. Routing Number
22. Insurance Policy Information		

Table 9: List of high risk nodes after running query for Email Address, Phone Number and Passport.

1. Password	2. Username	3. Name
4. Expiration Date	5. Address	6. Date of Birth
7. Social Security Number	8. Account Number	9. Credit Card Number
10. ID Card Information	11. Stolen Driver' s License Information	12. Bank Account Number
13. Monetary Amount	14. User Credentials	15. Signature
16. Employee Record	17. Fake ID Card Information	18. Account Information
19. Personal Identification Number (PIN)	20. Biographic Data	21. Check Information
22. Routing Number	23. Date	24. Insurance Policy Information

has the highest number (281) and National Identity Number keeps its impact restricted to 1 node, itself.

Not only have we considered the outdegree, we have also tabulated the indegree or number of parents for the identity nodes selected in Table 4. As per the data in Table 4, Social Security Number has the maximum number of parents, i.e. 27 which is almost thrice than indegree for any other node. This shows that breach of Social Security Number can be triggered by many more paths leading to it than other nodes in consideration. Social Security Number has the greatest number of parents hence it is at most risk of getting exposed. The more the number of parents, the more edges leading to it, and the higher the risk of it getting exposed given any parent is breached.

Table 10: List of high risk nodes after running query for Email Address, Phone Number and Driver's License.

1. Bank Account Information	2. Name	3. Patient Database
4. Address	5. Date of Birth	6. Social Security Number
7. Credit Card Number	8. Debit Card Information	9. ID Card Information
10. Monetary Amount	11. W-2 Form Information	12. Bank Card Expiration Date
13. CVV Code	14. Login Credentials	15. User Credentials
16. Signature	17. Employee Record	18. Fake ID Card Information
19. Account Information	20. Personal Identification Number (PIN)	21. Physical Address
22. Biographic Data	23. Passport Information	24. Customer Database
25. Check Information	26. Patient Medical Record	27. Date
28. Insurance Policy Information	29. Expiration Date	

Table 11: List of high risk nodes after running query for Email Address, Phone Number and National Identity Card.

1. Bank Account Information	2. Name	3. Address
4. Date of Birth	5. Social Security Number	6. Debit Card Information
7. ID Card Information	8. Stolen Driver's License Information	9. Monetary Amount
10. CVV Code	11. User Credentials	12. Employee Record
13. Fake ID Card Information	14. Account Information	15. Personal Identification Number (PIN)
16. Biographic Data	17. Passport Information	18. Check Information
19. Date	20. Insurance Policy Information	21. Expiration Date
22. Patient Database		

Table 12: Number of high risk nodes for query excluding those for query with evidence: Email Address and Phone Number.

Evidence Node Set	Unique High Risk Node Count
Email Address, Phone Number, SSN	8
Email Address, Phone Number, Passport	7
Email Address, Phone Number, Driver's License	14
Email Address, Phone Number, National Identity Card	8

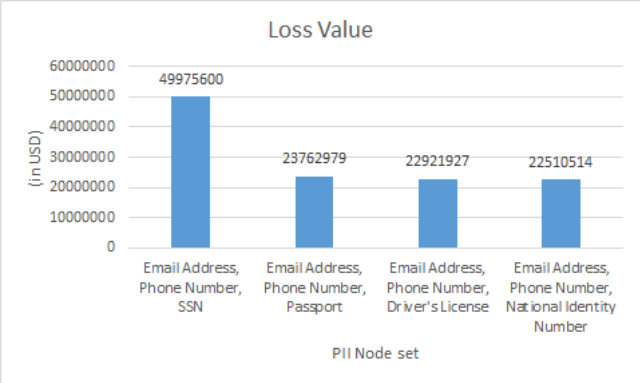

Figure 3: Loss Value of PII Node Set.

We also applied the query approach on our data in Identity Ecosystem. As we have seen in many blockchain-based IdM solutions, email address and phone number are used to set up accounts, hence we have noted them as the base case. Then we have listed down each of the set of PII exposed and run the query for that evidence set. We have listed the count of high and medium risk nodes after running the query of inferring the extent of breach when a set of identity nodes are exposed in Table 6. We observe the maximum number of high and medium risk nodes for Driver's License along with email address and phone number. This shows that using Driver's License for proving one's identity leads to more breaches than any other document and hence is most risky.

The list of all the high risk nodes after running the query of breach in Identity Ecosystem for various evidence sets has been captured in tables 7 to 11. Using these lists, we have captured the number of unique high risk nodes for each of the evidence set as compared to base case of email address and phone number in Table 12. This is the count of nodes which

are only present in their results but not in the results for the base case. We observe the count is highest for the evidence set of Driver's License which is almost double as compared to all the others. The result analysis is not very straightforward. These are the extra number of identity nodes which will be at risk of exposure if the PII email address, phone number and Driver's License are breached as compared to when only email address and phone number are exposed. The more the number of nodes in the result of this query, the more is the impact of breach. We would recommend to select an identity node with less number of nodes at risk.

Combining both the graph statistics and query approach, we also observed National Identity Card has the lowest prior probability for getting exposed and lowest number of children, hence it will be least at risk of exposing other identity nodes. Also, the number of parents for National Identity Number is the least putting it least at risk as compared to other PII nodes. The number of high and medium risk nodes getting breached after running query for National Identity Card as evidence set is the least confirming the observation from the first approach. Hence, using National Identity Card as identity verification document is recommended for blockchain-based solutions as it fulfills the authentication requirements and minimizes risk and liability.

5 CONCLUSION

In this paper, we studied the various PII used in identity verification for blockchain-based IdM services. We tried to find out ways to determine the best set of PII to be used for proving a person's identity in self-sovereign identity systems. Such PII can be used to complete authentication capabilities but also minimize risk and liability of exposure. We used the Identity Ecosystem, developed at the Center for Identity at the University of Texas at Austin, to provide two approaches. In the first approach, we investigated the PII graph node properties: their prior probability of risk and intrinsic loss value. For a specific PII node, its parents and children were observed to determine the likely exposure and influence it propagates to others. We conclude that the Social Security Number is the most risky node with highest prior probability and initial monetary loss and maximum number of children in the graph increasing its risk.

In the second approach, we leveraged the query of breach in the Ecosystem when a set of identity attributes are compromised. We studied the impact on other identity nodes by calculating the posterior probability and related loss. We also analyzed the differences in the nodes at risk in the highest risk level against the base case of documents required to open an IdM account, i.e., when only email address and phone number are exposed. The more the number of high and medium risk nodes in the results, the more the impact of the breach. We conclude the Driver's License has the highest impact and so using this kind of attribute for identity certification will increase risk to the individual's privacy.

Using both the graph statistics and query approach, we also concluded that National Identity Card is the most recommended identity document to use for verification in blockchain-based IdM solutions. It has the lowest prior probability, lowest number of children and number of high and medium risk nodes getting breached after running the query. Hence, using National Identity Card as identity verification document is a good option as it fulfills the authentication requirements and minimizes risk and liability.

The graph statistics and query approach give us different but justifiable results. The graph statistic approach analyzes the identity attribute nodes and their relationships with other nodes. The query approach based on the graph and Bayesian network adds to the probability model which is used in the Ecosystem and defines its results. As the number of identity theft and fraud stories continue to grow in our ITAP database, a more comprehensive study can be undertaken to capture a better picture of the blockchain-based identity verification.

6 FUTURE WORK

This paper helps the individual user by providing a guideline in choosing identity documents in a blockchain IdM with the help of Identity Ecosystem. In this work, individual loss tables were not calculated as the loss and risk of exposure in Identity Ecosystem are calculated by taking into account many identity theft stories covering such PII from different sources like media, government and victims. As a part of future work, we aim to provide more fine-grained recommendations to an individual about which PII document to choose by utilizing probability and risk loss *per* individual.

In this work, we evaluated the blockchain-based identity management solutions with respect to a standard framework used to predict risk associated with identity. We did not take into account the type of system used but the different type of PII involved in the IdM system. With the security and privacy increasing in the new blockchain-based IdM solutions, technologies can be customised to minimize breach of identity while performing authentication.

ACKNOWLEDGMENTS

This research has been supported in part by the The University of Texas Center for Identity Strategic Partners. The authors would like to thank James Zaiss for his help with the ITAP work and David Liau for his work on the latest implementation of Identity Ecosystem.

REFERENCES

[1] [n. d.]. 2018 End of Year Data Breach Report - Identity Theft Resource Center. https://www.idtheftcenter.org/

2018-end-of-year-data-breach-report/. (Accessed on 04/11/2019).

[2] [n. d.]. Civic - White Paper (draft)3-4.indd. https://tokensale.civic.com/CivicTokenSaleWhitePaper.pdf. (Accessed on 04/11/2019).

[3] [n. d.]. Civic Decentralized Reusable KYC Services - Blockchain-Powered. https://www.civic.com/solutions/kyc-services/. (Accessed on 04/11/2019).

[4] [n. d.]. Home | Identity Verification & KYC | Authenteq. https://authenteq.com/. (Accessed on 04/11/2019).

[5] [n. d.]. PII - Glossary | CSRC. https://csrc.nist.gov/glossary/term/PII. (Accessed on 04/11/2019).

[6] [n. d.]. ShoCard Identity Management Use Cases | ShoCard. https://shocard.com/identity-management-use-cases/. (Accessed on 04/11/2019).

[7] [n. d.]. A Zero-Knowledge Proof: Improving Privacy on a Blockchain | Altoros. https://www.altoros.com/blog/zero-knowledge-proof-improving-privacy-for-a-blockchain/. (Accessed on 04/11/2019).

[8] Christian Cachin and Marko Vukolić. 2017. Blockchain consensus protocols in the wild. *arXiv preprint arXiv:1707.01873* (2017).

[9] P. Dunphy and F. A. P. Petitcolas. 2018. A First Look at Identity Management Schemes on the Blockchain. *IEEE Security Privacy* 16, 4 (July 2018), 20–29. https://doi.org/10.1109/MSP.2018.3111247

[10] T. M. FernÃąndez-CaramÃľs and P. Fraga-Lamas. 2018. A Review on the Use of Blockchain for the Internet of Things. *IEEE Access* 6 (2018), 32979–33001. https://doi.org/10.1109/ACCESS.2018.2842685

[11] Marco Iansiti and Karim R Lakhani. 2017. The truth about blockchain. *Harvard Business Review* 95, 1 (2017), 118–127.

[12] R. Rana, R. N. Zaeem, and K. S. Barber. 2018. US-Centric vs. International Personally Identifiable Information: A Comparison Using the UT CID Identity Ecosystem. In *2018 International Carnahan Conference on Security Technology (ICCST)*. 1–5. https://doi.org/10.1109/CCST.2018.8585479

[13] Andrew Tobin and Drummond Reed. 2016. The inevitable rise of self-sovereign identity. *The Sovrin Foundation* 29 (2016).

[14] Razieh Nokhbeh Zaeem, Suratna Budalakoti, K Suzanne Barber, Muhibur Rasheed, and Chandrajit Bajaj. 2016. Predicting and explaining identity risk, exposure and cost using the ecosystem of identity attributes. In *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*. IEEE, 1–8.

[15] Razieh Nokhbeh Zaeem, Monisha Manoharan, and K Suzanne Barber. 2016. Risk kit: Highlighting vulnerable identity assets for specific age groups. In *2016 European Intelligence and Security Informatics Conference (EISIC)*. IEEE, 32–38.

[16] Razieh Nokhbeh Zaeem, Monisha Manoharan, Yongpeng Yang, and K Suzanne Barber. 2017. Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Computers & Security* 65 (2017), 50–63.

[17] Jim Zaiss, Razieh Nokhbeh Zaeem, and K Suzanne Barber. 2019. Identity Threat Assessment and Prediction. *Journal of Consumer Affairs* 53, 1 (2019), 58–70.