Privacy-Preserving Application of Blockchain Technologies

by
Brian Zabeti

A thesis submitted to the Department of Information and Logistics Technology,
College of Technology
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in Cybersecurity

Chair of Committee: Chris Bronk, Ph.D.

Committee Member: Yunpeng (Jack) Zhang, Ph.D.

Committee Member: Anthony Ambler, Ph.D.

University of Houston
April 2022

DEDICATION/EPIGRAPH

I dedicate this research to my daughter and the future generations, who may solve the problems we failed to solve in our time and make the world a safer and better place.

“Privacy is a right, not an opt-in/out checkbox.” – Me

“We make choices every day. Our choices can be part of a solution or a problem.” – Me

“Do not confuse motion and progress. A rocking horse keeps moving but does not make any progress.” – Alfred A. Montapert

ACKNOWLEDGMENTS

I appreciate my academic colleagues, committee, and faculty members for their assistance throughout my academic journey at the University of Houston. Dr. Arthur Conklin and Dr. Deniz Gurkan for sharing their valuable knowledge, and most significantly, Dr. Chris Bronk for his exceptional guidance and support.

I would like to thank my beloved wife and daughter for being supportive by creating an environment for me to dedicate my time to the graduate program and this research.

I would also like to thank my father, mother, and brother for their motivational support and high standards, helping me succeed. They have always wanted to see me get here and are proud of me now.

ABSTRACT

The blockchain rush and its rapid adoption in our daily life raise new questions and concerns regarding user and data privacy; An important topic that is typically overlooked due to the lack of proper standards and procedures in an under-regulated technology. Identity theft and data breaches are already impacting users on the internet since enterprises are collecting, processing, and storing more consumers' data in their databases, making them valuable and interesting targets for adversaries and criminal hackers. In most of the existing public blockchains, all transactions are visible in a public ledger where its data can be accessed and processed to profile users for marketing and other purposes without the user's consent and knowledge. Such enterprises' interest and influence in blockchain networks and considering their centralized nature and track record can raise the question of data ownership and privacy once again. Blockchain technologies and their decentralized nature are meant to shift the power by giving more control and freedom to users and individuals. But without proper privacy and security controls to protect their data, this could be a false optimism and create a bigger issue instead of solving one. This research aims to contribute, review, and highlight possible solutions for using privacy-focused blockchain technologies and their latest developments to address current and upcoming user and data privacy challenges. Additionally, a Blockchain Reference Model is introduced, which may be used as a reference to standardize the blockchain development.

Keywords: Privacy-preserving blockchain, Blockchain applications, Blockchain technologies, Decentralized Identity Management, Self-Sovereign Identity

TABLE OF CONTENTS

LIST OF TABLES

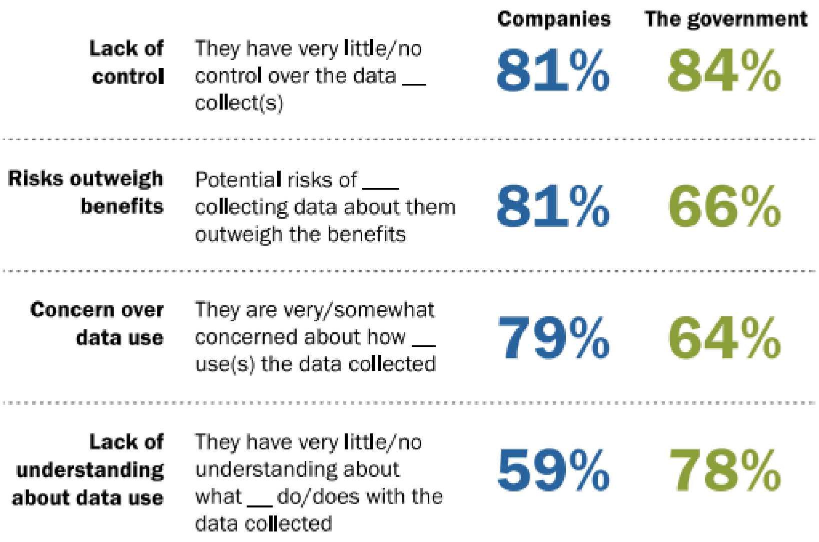
LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Problem Statement

A recent Federal Trade Commission (FTC) [1] report shows an alarming increase in identity theft and fraud in 2021 over the previous year. The consumers reported losing more than $5.8 billion, a 70 percent increase over the previous year [1]–[3]. A 2019 research survey by Pew research center shows that six out of ten American believe that it is not possible to go through their daily life without their data being collected by the government or companies [4]. As shown in Figure 1, research also shows the majority of Americans feel they have little or no control over their data collected by these entities [4].



		Companies	The government
Lack of control	They have very little/no control over the data __ collect(s)	81%	84%
Risks outweigh benefits	Potential risks of ___ collecting data about them outweigh the benefits	81%	66%
Concern over data use	They are very/somewhat concerned about how __ use(s) the data collected	79%	64%
Lack of understanding about data use	They have very little/no understanding about what __ do/does with the data collected	59%	78%

Figure 1. Data collection and privacy concerns survey [4]

The Identity Theft Resource Center (ITRC) [5] 2021 annual report shows an alarming 68 percent increase in data compromises over the previous year which is 23 percent over the previous all-time high [6]. The top 10 breached data attributes in 2021 for Personally Identifiable Information (PII) include: Names, Full Social Security Number (SSN), Date of Birth (DOB), Current home address, Medical history/condition/treatment/diagnosis, Driver's license/State ID, Bank account number, Medical insurance account number, Phone number, and payment card full number [6]. The report also mentions the top compromised data types from breaches between 2017 to 2021 in order are: social security numbers, personal health information, driver's license, and bank account [6] presented in Figure 2.

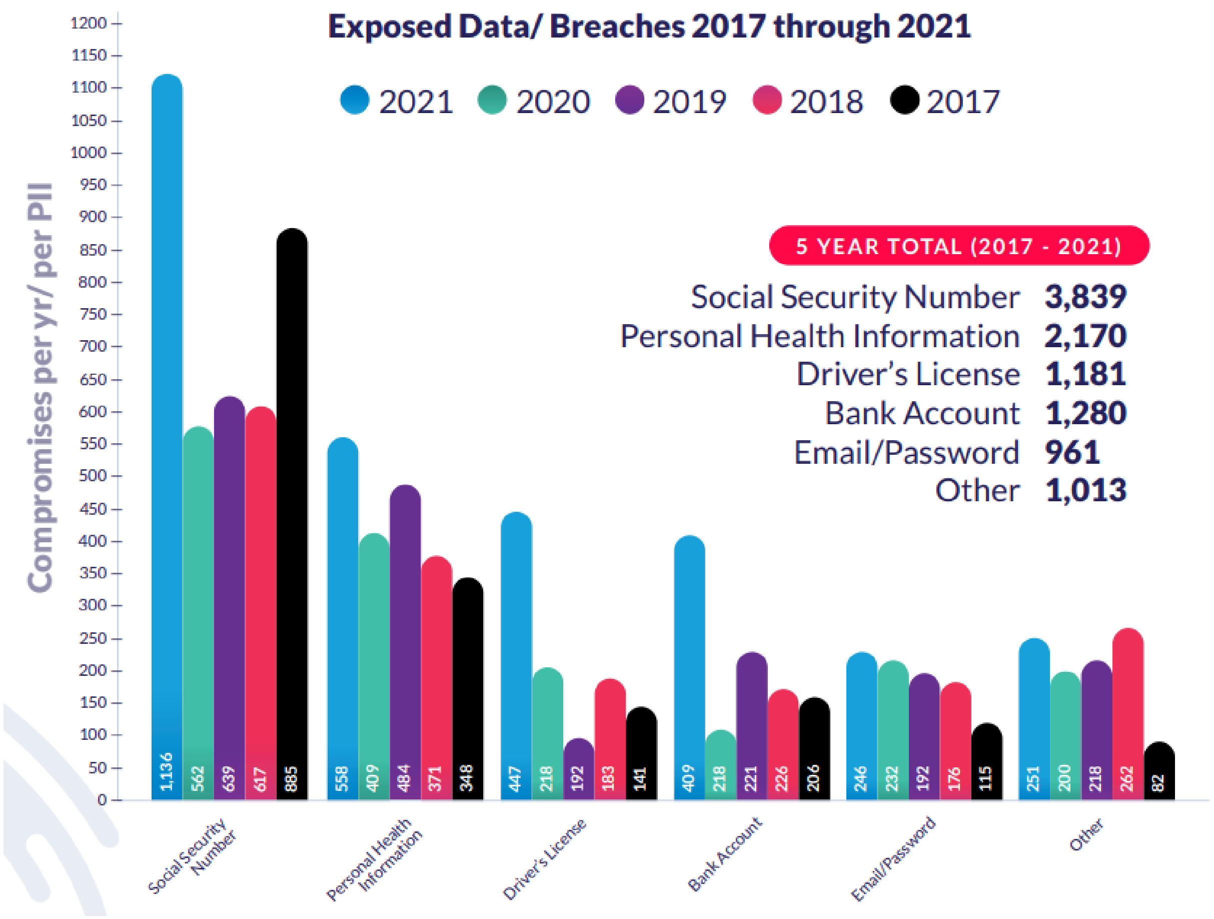


Figure 2. Compromised data types between 2017-2021 [6]

These stats are based on reported and known breaches; many private and smaller organizations still fail to report or even detect breaches inside their organizations. The issue of user/data privacy, safeguarding costs, risk and liabilities affects users, companies, the government, and everyone involved.

Research Questions

This research intended to answer the following questions:

- How to better protect PIIs and improve user and data privacy?
    - Can it be protected against data leaks and breaches?
    - How to give users the true power to opt in/out of marketing and advertisement?
    - Is there a way to stop unnecessary data sharing?
    - Is there a better alternative to centralized user data storage?
    - Can we stop handing out the actual data at the door (sign-up) and adhere to regulations such as KYC [7]?
- Can Blockchain technologies solve and/or improve privacy challenges?

Research Goals

This research aims to contribute, review, and highlight possible solutions for using privacy-focused blockchain technologies and their latest developments to address current and upcoming user and data privacy challenges.

CHAPTER 2

BLOCKCHAIN

Overview

Blockchain technologies have quickly become important in many industries, to the extent that it is no longer the question “if” the global mass adoption will happen but “when” it will happen. Almost every industry is either integrating their existing eco-system or finding a practical use for their applications and addressing the relative challenges. Government agencies, Tech, Financial, Logistics, Healthcare, and Insurance industries, are the early adopters in this race.

The concept of blockchain technology has been around since 1991 as Herber and Stornetta for the first time described a cryptographically secured chain of blocks in *Secure names for bit-strings* paper [8]. But it wasn’t until 2009 when Satoshi Nakamoto published the famous Bitcoin whitepaper that started a technological innovation [9] as shown in Figure 3. Bitcoin introduced a decentralized network in which participants could use its digital currency as a medium of exchange without an intermediary to conduct a financial transaction. Although, this was a remarkable advancement, in 2013, Vitalik Buterin introduced a programmable network that offered more than just a money exchange [10] [11]. Development of Ethereum resulted in new features such as Smart Contracts, Decentralized Finance (DeFi), and Non-Fungible Token (NFT) [11]. Gavin Wood, the co-founder of Ethereum, developed a high-level programming language called Solidity, enabling developers to create their own smart contracts on the Ethereum network opening the door for more possibilities [11].

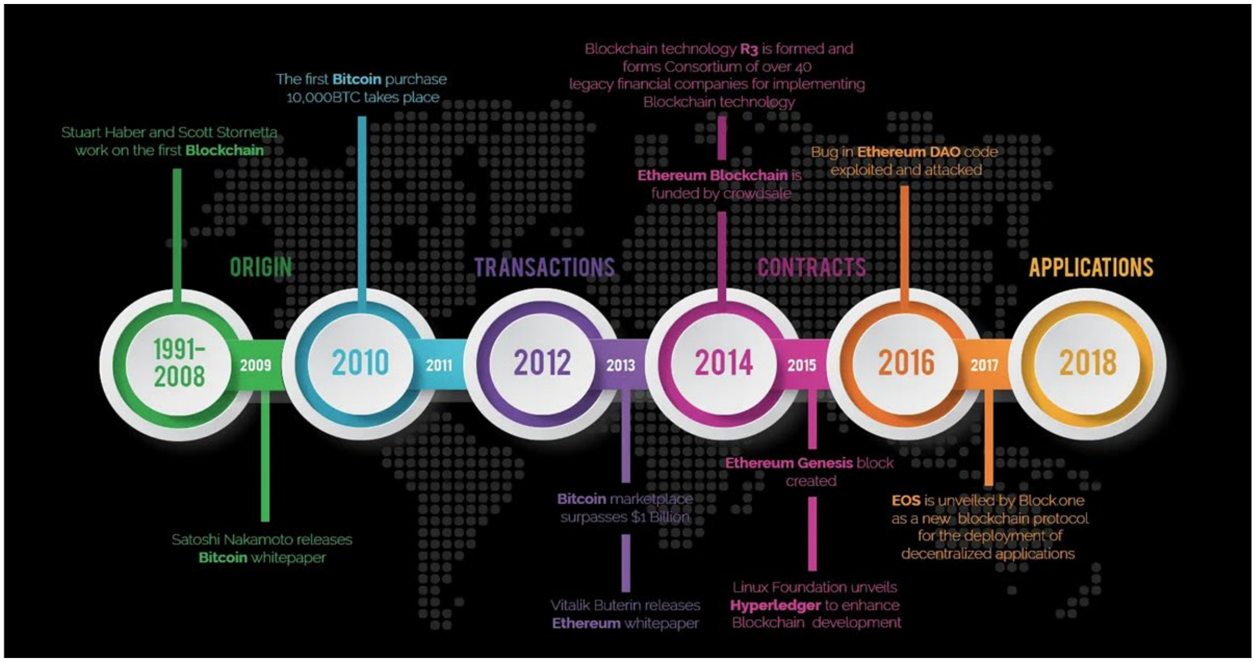

Stuart Haber and Scott Stornetta work on the first Blockchain
The first Bitcoin purchase 10,000BTC takes place
Blockchain technology R3 is formed and forms Consortium of over 40 legacy financial companies for implementing Blockchain technology
Ethereum Blockchain is funded by crowdsale
Bug in Ethereum DAO code exploited and attacked
ORIGIN
TRANSACTIONS
CONTRACTS
APPLICATIONS
1991-2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Satoshi Nakamoto releases Bitcoin whitepaper
Bitcoin marketplace surpasses $1 Billion
Vitalik Buterin releases Ethereum whitepaper
Ethereum Genesis block created
Linux Foundation unveils Hyperledger to enhance Blockchain development
EOS is unveiled by Blockone as a new blockchain protocol for the deployment of decentralized applications

Figure 3. Blockchain history timeline [12]

Blockchain technologies are relatively new and still have many potentials and challenges which can only be discovered over time as it gains more popularity and people find more use-cases in real-world applications. Despite the ongoing cryptocurrency rush since 2020, many are still unaware of the underlying blockchain technologies and the possible solutions that can be offered to today's existing challenges. The current focus is mainly on blockchain's financial applications, but the technology can be used for more than that, such as in supply chains, healthcare, privacy, security, voting, election integrity, and more. Before applying blockchain technologies in any industry, it requires an understanding of what is under its hood.

What is it?

Blockchain is a decentralized, peer-to-peer (P2P) recording information system featuring a tamper-resistant and tamper-evident distributed digital ledger for transactions on its network. This is achieved as all transactions are duplicated and then distributed across the participating nodes in the network to first verify and then store them. Transactions are stored in blocks that are cryptographically linked together, where each block's header contains the hash of its previous block's header, forming the "Blockchain" as shown in Figure 4.

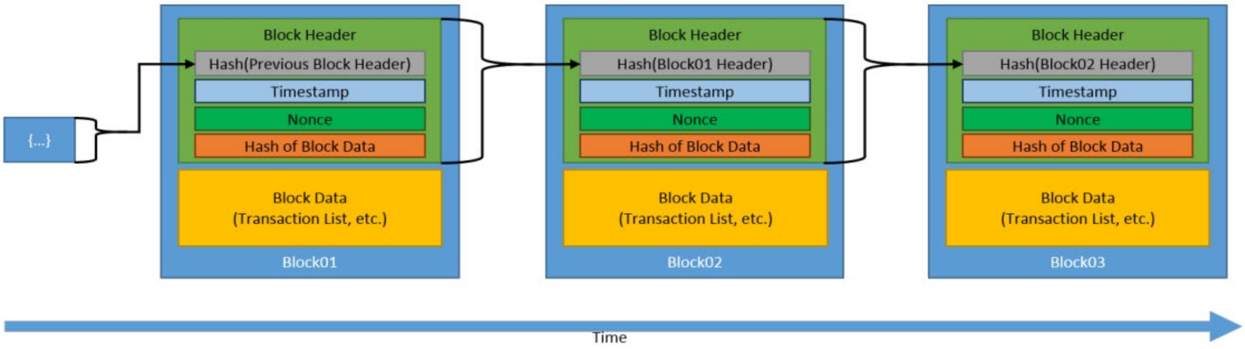


Figure 4. Generic chain of blocks [13]

Blockchain Components

The main components that make up a blockchain network are Transactions, Blocks, Ledgers, and a combination of different cryptography types and algorithms. Each blockchain implementation may use a different cryptographic algorithm or even additional components for its architecture, but the main components are always present.

Cryptography

Cryptography types are grouped into three main categories, Symmetric-Key Cryptography, Asymmetric-Key Cryptography, and Hash Functions. Out of the three categories, only two are commonly applied to blockchain architecture.

Asymmetric-Key Cryptography. (also known as Public-Key Cryptography) employs a pair of mathematically associated keys named public and private keys to verify the authenticity and integrity of a transaction or message. In a blockchain network, asymmetric-key cryptography provides a mechanism to [13]:

- Verify asset ownership (using Private keys)
- Digitally sign transactions (using Private keys)

- Derive addresses (using Public keys)
- Signature verification (using Public keys)

Hash functions. Hash functions are preimage-resistant (one-way) cryptographic operations. Such an operation uses a cryptographic hash function also known as “hashing,” to derive a unique output called “digest” from given data input. Meaning any change to the input data will result in an entirely different digest. (Note: There are some weak hashing algorithms that are considered insecure). Blockchain networks benefit from hash functions as they are quicker and computationally lighter than key-based cryptography operations. Hashing is used to establish the correct link between blocks (specifically block headers) in addition to ensuring the integrity of stored data in each block of the blockchain.

Transactions

A transaction in a blockchain network represents an interaction between parties [13]. Each transaction at least contains information such as a medium of exchange (digital asset), amount, source address, destination address, and proof of access/ownership to the referenced digital asset. Typically, the ownership is proved by digitally signing the transaction using the associated private key.

Blocks

Blocks are technically datasets that are linked together to form a blockchain. Each block is broken into two main parts: a block header and block data. (see Figure 4)

Block header. Contains the metadata such as the timestamp, block size, hash of the block data, and hash of the previous block’s header.

Block data. Contains a list of transactions and other data.

Ledgers

A ledger is a distributed collection of digital transactions and/or events. In a blockchain network, the ledger acts as a distributed database in which every network participant (node) holds a local copy of it. Technically, the blockchain itself is considered a digital ledger with a sequence of blocks.

How does it work?

The process can be shortened to three main processes: request, validation, and block creation. At first, someone requests a transaction that is broadcasted to the nodes in the network. The network then tries to validate the transaction and requester's status. Once validated, the transaction is queued until a node (publishing node), publishes a new block and appends it to the network. The published block is permanent and cannot be altered.

Categories

Blockchain networks at the highest level are categorized by their network access level (Public and Private) or participation model (Permissionless and Permissioned). The difference between them depends on who can participate in the network and their role as a participant. But a more detailed blockchain categorization divides the network architecture into four categories as shown in Figure 5.

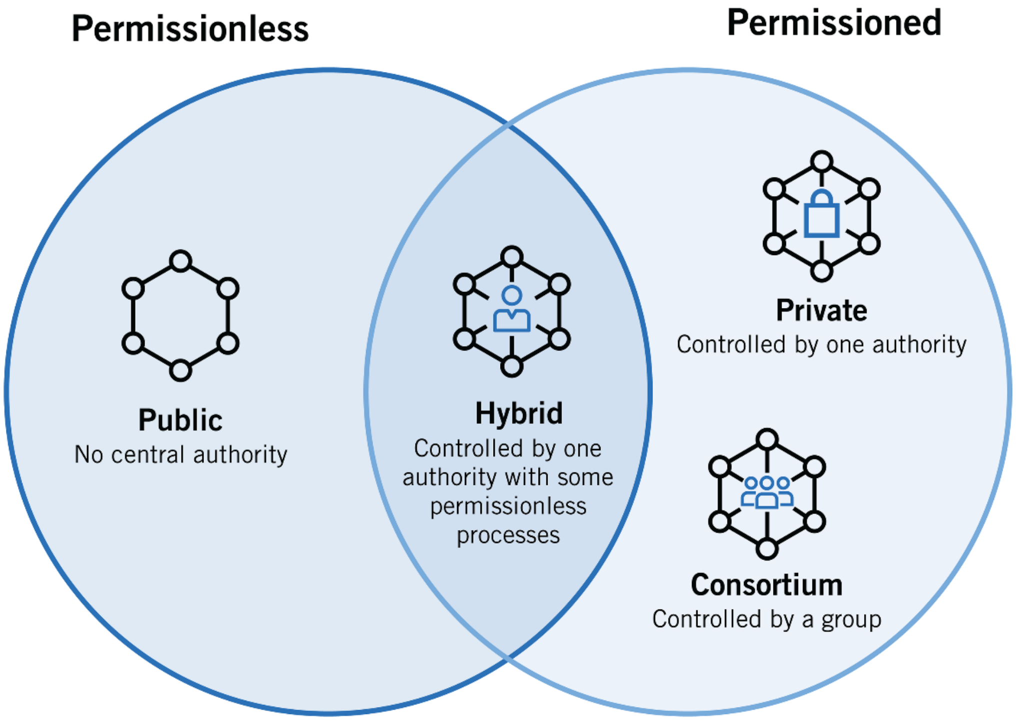


Figure 5. Blockchain Categories [14]

- Public
    - Participation Model: Permissionless
    - Structure: Decentralized with no central authority
    - Anyone can join the network as a node and validate transactions
- Private
    - Participation Model: Permissioned
    - Structure: Centralized and controlled by one authority
    - Restricted access, limited and defined by an entity (Business, government agency, etc.)
- Consortium
    - Participation Model: Permissioned
    - Structure: Centralized and controlled by a group
- Hybrid
    - Participation Model: Hybrid
    - Structure: Centralized and controlled by one authority (or a group) with some permissionless processes

Consensus Mechanisms

In a blockchain network, consensus refers to the protocol and methods enacted to achieve an agreement between the network participants. There are many consensus models such as Proof of Work (PoW), Proof of Stake (PoS), Delegated PoS, Proof of Authority/Proof of Identity, Proof of Elapsed Time, Round-Robin and more.

Proof of Work and Proof of Stake are the most common protocols used in today's blockchain networks. They both are Probabilistic finality meaning when a transaction sinks deeper into the chain; it becomes more difficult to revert it. The fault tolerance in both consensus protocols is 51% which means if someone manages to obtain 51% of computational power (in PoW based network) or coins (in PoS based network) they can create a long private chain to replace a valid chain also known as double-spending [15].

Proof of Work

In a Proof of Work (PoW) consensus (see Figure 6), miners provide computational power to the network. Anyone that meets the hardware requirements can start mining. They compete at each round to solve a cryptographic puzzle and the winner gets to publish the new block and receives a reward. PoW requires a lot of computational power as miners need to keep adjusting the value of nonce (an arbitrary number used once in hashing processes) in order to get the correct answer. Examples of PoW consensus networks are Bitcoin and Monero [15].

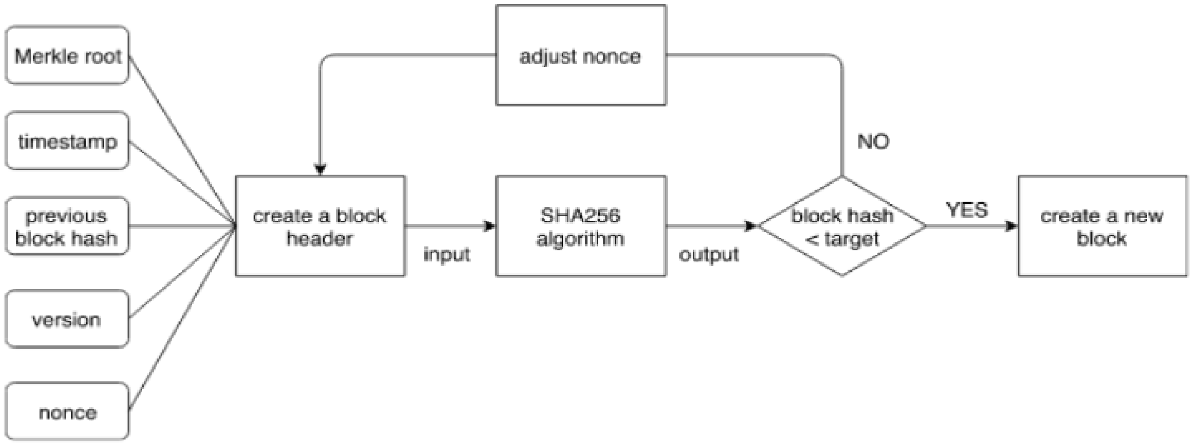


Figure 6. Flow of PoW [15]

Proof of Stake

In Proof of Stake (PoS) consensus (see Figure 7), validators provide computation power to the network. Validators require to stake and hold a minimum of coin amount on the network. The consensus is achieved by validators on the network. The more coins they hold, the more authority they have. PoS is more energy efficient since there is no computational race to adjust the nonce. Examples of PoS consensus networks are Solana and Cardano.

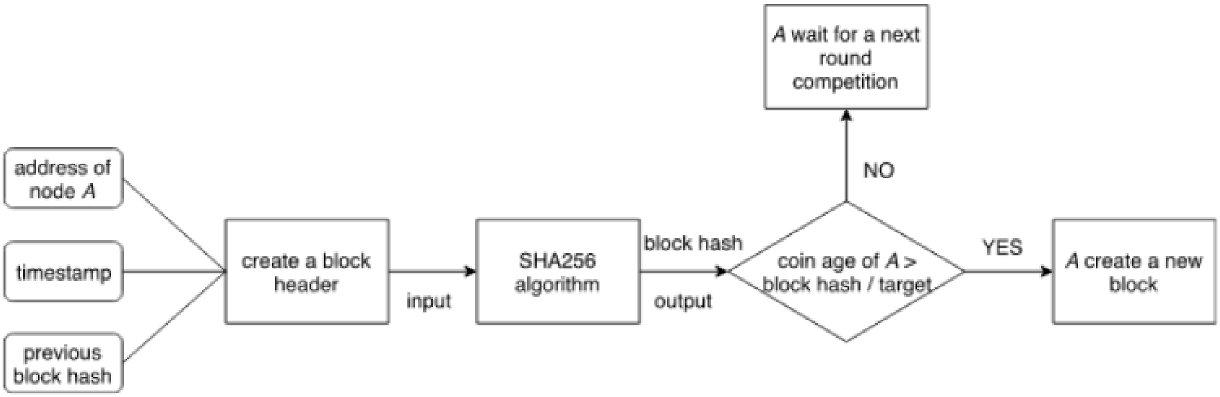


Figure 7. Flow of PoS [15]

Smart Contracts

It is a digital contract containing a collection of code and data that is digitally signed and deployed on the blockchain network to satisfy common contractual conditions. Smart contracts can serve many functions such as storing data, performing calculations, automations, and more to extend and enhance the blockchain network. Smart contracts leverage the existing blockchain network resources and can have their own tokens based on the parent blockchain's coin [13].

Blockchain Reference Model

Blockchain's technology stack may differ from one network to another. Lack of standardization and purpose-specific network architecture used in developing each blockchain could introduce problems such as incompatibility. Similar to the TCP/IP stack and OSI model, an standard reference model can be used in developing blockchains to ensure higher cross-chain compatibility and efficiency. This research proposes the following six layers architecture shown in Table 1.

Layer	Name	Function	Example
5	**Application / Presentation**	Acting as the User Interface (UI) for business logic and user interactions	• UI • Decentralized Applications (dApps) • Application integration • Programming Languages
4	**Service**	Services to enable application operations and connection to other technologies.	• Smart Contracts • Wallets • Digital Assets • Digital IDs • Connectors
3	**Network / Transportation**	Defining the network participation requirement, base protocol, and method of consensus. Acting as a transportation interface for the P2P network which decides how data is controlled and transferred.	• Permissionless/Permissioned • Consensus mechanism • Sidechains
2	**Data**	Lists of chained blocks (Linked list) and pointers referring to block positions. Cryptographic algorithm and the binary tree of hashes also known as Merkle tree.	• Digital signatures • Hash • Transactions
1	**Infrastructure**	Blockchain as a Service (BaaS) to control the nodes. Solana [16], Ethereum [17], Polkadot [18], and Avalanche [19] networks are three examples of layer one blockchains.	• Compute • Storage • Virtual Machines (VMs)
0	**Foundation / Physical**	The foundational physical components that enable blockchain technologies as a whole to exist and function.	• Hardware • Internet

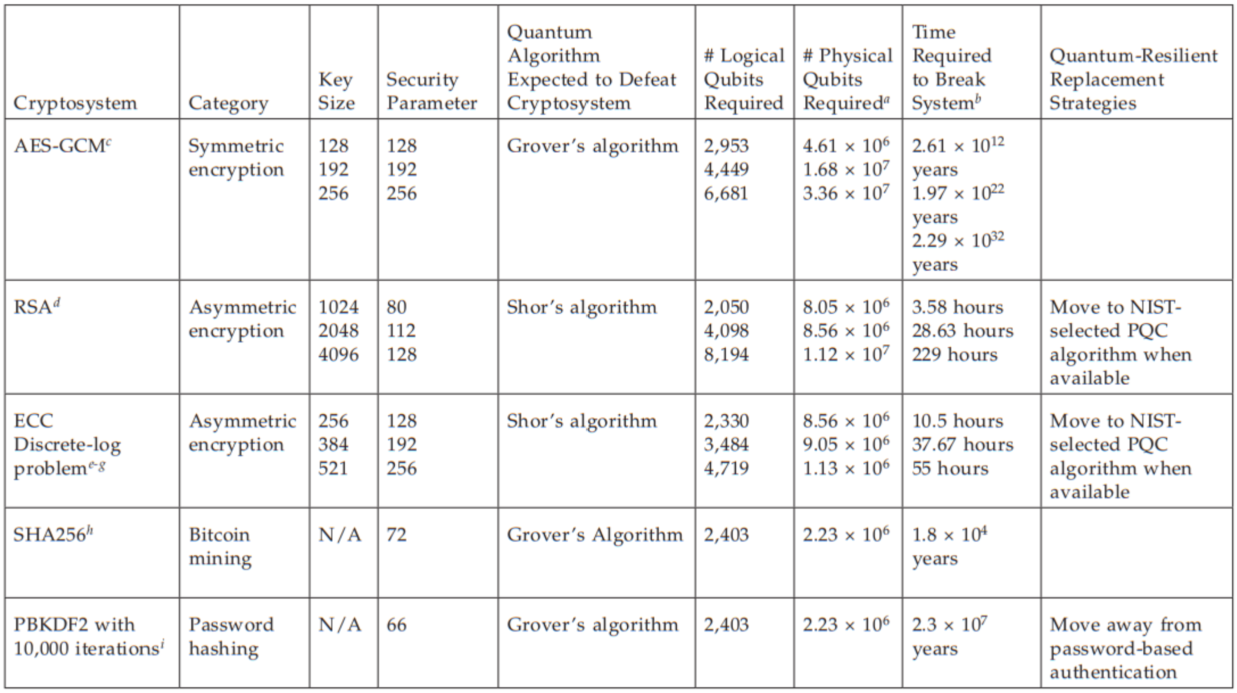
Table 1. Blockchain Reference Model

Security Challenges

Blockchain is still considered an emerging technology; therefore, security holes and vulnerabilities can be discovered over time. Some of the known security challenges are [20]:

- Phishing attacks
    - Attempt to steal user's credentials and private keys through email and/or social engineering tactics
- Routing attacks
    - Man-in-the-middle (MITM) and sniffing the traffic to intercept sensitive data.
- Sybil attacks
    - Using a large number of false identities to flood and crash the system.
- Majority attack (51% attack)
    - Gaining more than 50% control of a network to create a separate chain, also known as double-spending attack

Another security threat to blockchains and specifically their cryptographic algorithms is Quantum computing. Although today's quantum computers are not as powerful and still far from being practical and adopted, the threat posed by their potential computational power supremacy cannot be easily ignored. The National Academics of Science, Engineering, Medicine conducted a consensus study of quantum computing and its implications for cryptography. The result shows that asymmetric cryptographic algorithms based on the discrete log problem and integer factorization are most vulnerable to the current quantum algorithms, specifically Shor's algorithm. Some examples are RSA, Elliptic Curve (EC), and key exchange protocols based on the Diffie-Hellman (DH). Theoretically, a quantum computer with around 2,300 logical qubits can break RSA 1024 in less than a day [21] see Table 2.



Cryptosystem	Category	Key Size	Security Parameter	Quantum Algorithm Expected to Defeat Cryptosystem	# Logical Qubits Required	# Physical Qubits Required[a]	Time Required to Break System[b]	Quantum-Resilient Replacement Strategies
AES-GCM[c]	Symmetric encryption	128 192 256	128 192 256	Grover's algorithm	2,953 4,449 6,681	4.61×10^{6} 1.68×10^{7} 3.36×10^{7}	2.61×10^{12} years 1.97×10^{22} years 2.29×10^{32} years	
RSA[d]	Asymmetric encryption	1024 2048 4096	80 112 128	Shor's algorithm	2,050 4,098 8,194	8.05×10^{6} 8.56×10^{6} 1.12×10^{7}	3.58 hours 28.63 hours 229 hours	Move to NIST-selected PQC algorithm when available
ECC Discrete-log problem[e-g]	Asymmetric encryption	256 384 521	128 192 256	Shor's algorithm	2,330 3,484 4,719	8.56×10^{6} 9.05×10^{6} 1.13×10^{6}	10.5 hours 37.67 hours 55 hours	Move to NIST-selected PQC algorithm when available
SHA256[h]	Bitcoin mining	N/A	72	Grover's Algorithm	2,403	2.23×10^{6}	1.8×10^{4} years	
PBKDF2 with 10,000 iterations[i]	Password hashing	N/A	66	Grover's algorithm	2,403	2.23×10^{6}	2.3×10^{7} years	Move away from password-based authentication

Table 2. Quantum Resiliency of Current Cryptosystems [21]

CHAPTER 3

PRIVACY-PRESERVING APPLICATION

Overview

In the Big Data era, many organizations, both public and private collect, store, and utilize personal data. Bigger companies process and store even more data in their data warehouses to offer various products and services for a personalized experience. Therefore, there is no surprise in seeing increased identity theft and data breaches. Cybercriminals regularly target organizations to steal user data and personal information to sell them on the dark web or use them directly for malicious intentions, compromising users' privacy.

This research asks if blockchain technologies can offer a solution to fix the current data privacy challenges. Using blockchain technologies and techniques to achieve a decentralized solution where users are in control of their own data while information is not stored with third parties, service providers, or identity providers may be the answer to solve some of these privacy challenges. This section focuses on reviewing the main technologies and techniques in blockchain that can be used to enhance user and data privacy.

What is Privacy?

"Privacy is one of the most important concepts of our time, yet it is also one of the most elusive." [22] The ever-growing technology increases the data and information availability. Almost everyone, from scholars, activists, to policymakers struggling to define privacy. Some may even call it an impossible task [22]. Many businesses collect users' data to improve their services and products offered to consumers, but this information is often shared with third parties for multiple reasons, from performance optimization, deep profiling, targeted advertising, and in-depth analysis. Whether they are shared or not, collecting users' data can impose privacy risks and costs to the businesses responsible for safeguarding the data and, most importantly, the users themselves.

The 2019 *Comprehensive Survey on Big Data Privacy Protection* [23] published in IEEE journal noted, "User privacy has evolved into a critical issue in various data mining operations. User privacy has turned out to be a foremost criterion for allowing the transfer of confidential information. The intense surge in storing the personal data of customers (i.e., big data) has resulted in a new research area, which is referred to as privacy-preserving data mining (PPDM)." [23]

Social Media, Big Data, Internet of Things (IoT)

The amount of online data has been rapidly growing since the rise of Web 2.0, with the total amount of online data is estimated to be around 100 zettabytes, or 100 billion terabytes. Clive Humby, a British mathematician and Tesco marketing mastermind, said, "Data is the new oil" in 2006 [24] see Figure 9. But the quote and idea did not get much public attention until the Economist published an article named "The world's most valuable resource is no longer oil but data" in May of 2017 [25]. Humby's famous quote has been challenged by many since data is not a finite resource and can be reused or replicated without it being destroyed or declining in quality; instead, data can be more accurately referred to as an asset or investment [24].

Figure 8. The world's most valuable resource is no longer oil but data [25]

Technologies and Techniques

A summary of current privacy-based techniques is presented in Figure 10. The key elements referenced in Figure 10 are further analyzed in detail to review their applications in user and data privacy. Some of these elements have been around for many years and may have been integrated into existing technologies such as mixing, a common technique used in email security solutions. But some of these elements are still a concept or a work in progress, such as zkSTARK.

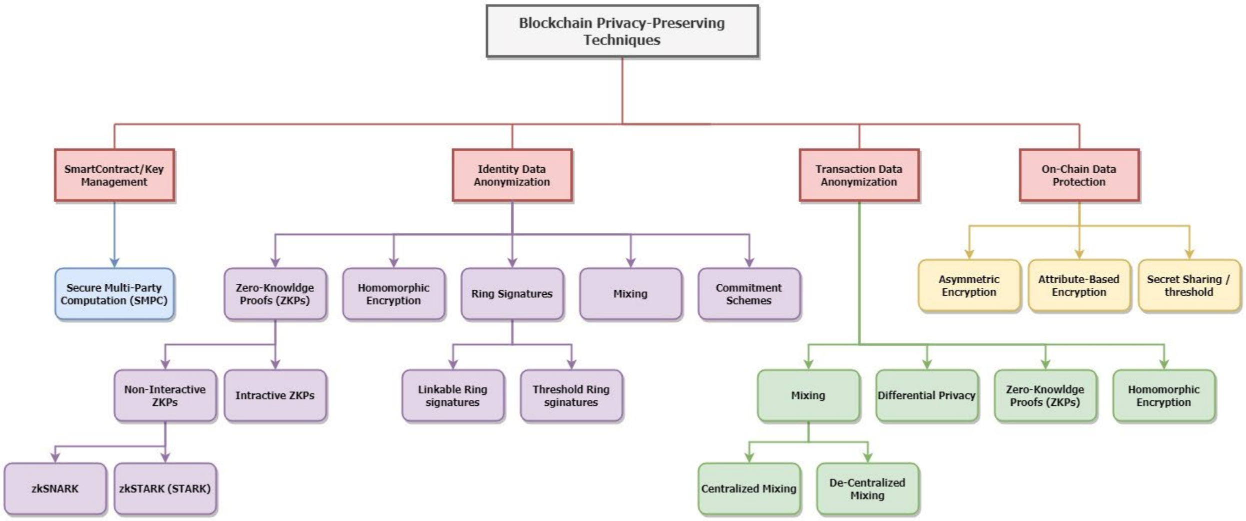

Blockchain Privacy-Preserving Techniques
SmartContract/Key Management
Identity Data Anonymization
Transaction Data Anonymization
On-Chain Data Protection
Secure Multi-Party Computation (SMPC)
Zero-Knowldge Proofs (ZKPs)
Homomorphic Encryption
Ring Signatures
Mixing
Commitment Schemes
Asymmetric Encryption
Attribute-Based Encryption
Secret Sharing / threshold
Non-Interactive ZKPs
Intractive ZKPs
Linkable Ring signatures
Threshold Ring sginatures
Mixing
Differential Privacy
Zero-Knowldge Proofs (ZKPs)
Homomorphic Encryption
zkSNARK
zkSTARK (STARK)
Centralized Mixing
De-Centralized Mixing

Figure 9. Taxonomy of Privacy-preserving techniques for blockchain [26]

Secure Multi-Party Computation (SMPC)

One of the biggest privacy main challenges is utilizing private data without revealing it [27]. SMPC aims to address this by splitting data between a known number of parties where it requires the majority's cooperation to perform the distributed computation of given inputs to generate the output and reveling the data. SMPC can be used to address account and key management issues without requiring any third parties. [26]

Zero-Knowledge Proof (ZKP)

The history of ZKP can be traced back to the 1980s, when interactive ZKP was initially introduced in *The Knowledge Complexity of Interactive Proof-Systems* by MIT scientists Shafi Goldwasser, Silvio Micali, and Charles Rackoff [28]. ZKP is a cryptographic protocol that enables two parties, the "verifier" and "prover" to prove a given statement is true based on mathematical probability without revealing the secret to the verifier. It aims to address three properties:

- **Completeness:** If the prover's statement is rendered true, then the prover can always carry out a successful proof.

- **Soundness:** If the statement is proved to be false, a cheating prover cannot convince the verifier that the statement is true, except for a small probability “statistical soundness”.
- **Zero-Knowledge:** Besides verifying the True or False state of the given statement, Verifier or eavesdroppers won’t be able to obtain extra information from the transcript.

The interactive nature of the ZKP protocol limits its practical applications but using the random oracle model based on the Fiat-Shamir heuristic can address this issue [26]. An implementation example of this can be seen in Zerocoin cryptocurrency [29] developed by Matthew Green, a Johns Hopkins professor, and his fellow students which is now integrated into Zcash cryptocurrency [30].

zkSNARK

The acronym stands for Zero-Knowledge Succinct Non-Interactive Argument of Knowledge. An enhanced and practical alternative to ZKP which offers more possibilities for its application. Similar to ZKP, in zkSNARK prover can prove possession of information without revealing the information in a non-interactive mechanism between prover and verifier. zkSNARK properties are embedded in the acronym as follow:

- **Zero-Knowledge:** The verifier can only see the validity of the statement, whether it is True or False, and nothing about the proof.
- **Succinct:** With a relatively small proof length bytes, verifications are achieved in a few milliseconds.
- **Non-Interactive:** The only information needed is the proof from the prover, which any verifier can validate. This is an important function for privacy-preserving applications.
- **Arguments:** Similar to “proofs” in ZKP but with the prover being limited in polynomial time that can only achieve “computational soundness”. They also make it possible to use cryptographic primitives. [31]
- **Knowledge:** Requiring a witness or secret value before enabling the argument constructions by the prover.

Utilization of zkSNARK enables a blockchain to provide a proof that X calculation was done correctly without revealing its parameters; Therefore, a trust can be established. Currently, zkSNARK requires a trusted setup phase in which a common reference string (CSR), also known as public parameters. CSR is a random cryptographic value generated between the verifier and prover and then deleted. If the secret randomness used to generate the CSR is compromised, then false proof can be created that looks valid to the verifier. Zcash [30], [32] addresses this issue with Multi-party Computation Ceremonies (MPC); a protocol proposed by Sean Bowe [33] that allows collaborative CSR generation by multiple independent parties which can only be compromised if all participants are compromised or dishonest [30], [32]. zkSNARK has other disadvantages, such as being taxing on blockchain resources and its utilization of elliptic curves, a quantum-susceptible cryptography as previously discussed in CHAPTER 2 .

zkSTARK

A proof of concept and derivation of zkSNARK proposed by Eli Ben-Sasson [34] that can address many of the deficiencies in its predecessor in addition to using quantum-resistant hash functions. The acronym stands for Zero-Knowledge Scalable Transparent ARgument of Knowledge in which its properties are also embedded into as:

- **Zero-Knowledge:** Similar to zkSNARK, offering input privacy to prover without exposing its input to the blockchain.
- **Scalable:** Off-chain based computation that reduces verification costs.
- **Transparent:** Defined as “public randomness”; it means the randomness used in all verifier’s messages and queries are public without the need for a trusted setup phase.
- **Argument of knowledge:** A proof that ensures only the valid prover executed the computation and its required inputs.

One of the zkSTARK disadvantages is its 1000 times longer proof lengths over its predecessor, which may require further research to improve. Another one is the possible participant exclusion due to limited off-chain computation resources. Although it may take time to fully test and improve zkSTARK applications, if the concept is proven, it could be used to enhance privacy and address the existing trust and zero-knowledge debate.

Fully Homomorphic Encryption (FHE)

Simply put, Homomorphism refers to preserving the relation to X after transforming X to Y. In cryptography, this technique enables processing of encrypted data or ciphertext without decrypting it. In the late 1970s, The Homomorphic Encryption, also called Privacy Homomorphisms, was first introduced by scientists from MIT [35]. In 2009, IBM researcher Craig Gentry introduced the FHE [36] to complement the initial research, but there has not been a practical application for it due to its high computing power requirements. With the hardware and software advancements and increased computational power in recent years, the concept is gaining more popularity since IBM is now offering it as a service to their clients [37]. FHE application in privacy-preserving blockchain can provide a mechanism to process encrypted user data without the need to decrypt it. In addition, the stored information is encrypted in a way that can eliminate the risk of personal data exposure in a data breach situation.

Mixing

Mixing is an anonymization technique in which a bulk of data from multiple users are delayed and then forwarded at the same time in random order. In blockchain, mixing technique can be considered as a proxy service which can be used to un-correlates the wallet addresses in the transaction’s history.

Ring Signatures

In 2001 Ronald L. Rivest1, Adi Shamir2, and Yael Tauman wrote the paper “How to Leak a Secret” [38], introducing the ring signatures scheme. The research was based on the initial work of David Chaum and Eugene van Heyst for group signature scheme in 1991 [39].

With ring signatures, it is possible to indicate a set of possible signers without revealing which member produced the signature; therefore, offering signer-ambiguity. Any user can choose any set of possible signers that includes himself and sign any message by using his secret key and the others' public keys without getting their approval or assistance. Monero is a privacy-preserving blockchain that uses ring signatures and ring Confidential Transactions (CT) to maximize users' privacy [40], [41].

Decentralized IDentifier (DID)

DIDs are a new type of identifier that enables verifiable, decentralized digital identity. A DID refers to any subject (e.g., a person, organization, thing, data model, abstract entity, etc.) as determined by the controller of the DID. The controller of a DID can prove control over it without requiring permission from any other party. DIDs are URIs that associate a DID subject with a DID document allowing trustable interactions associated with that subject. [42]

A DID is a simple text string consisting of three parts, as shown in Figure 11:

1. The DID URI scheme identifier
2. The identifier for the DID method
3. The DID method-specific identifier.

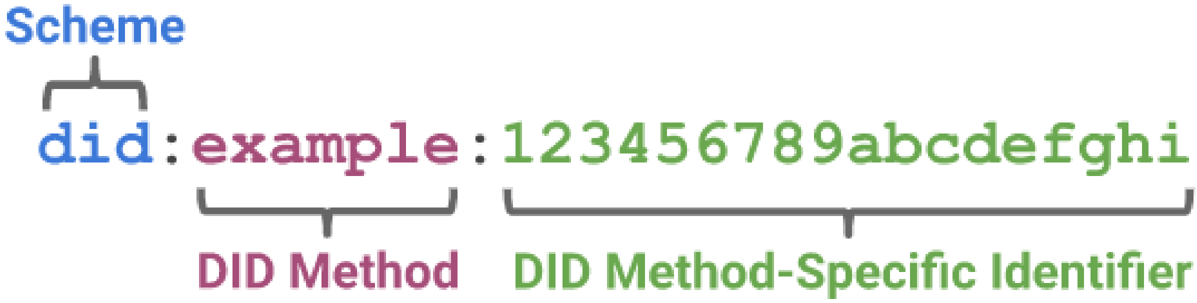

Scheme
did:example:123456789abcdefghi
DID Method
DID Method-Specific Identifier

Figure 10. A simple example of a decentralized identifier (DID) [42]

Decentralized Self-Sovereign Identity (DSSI)

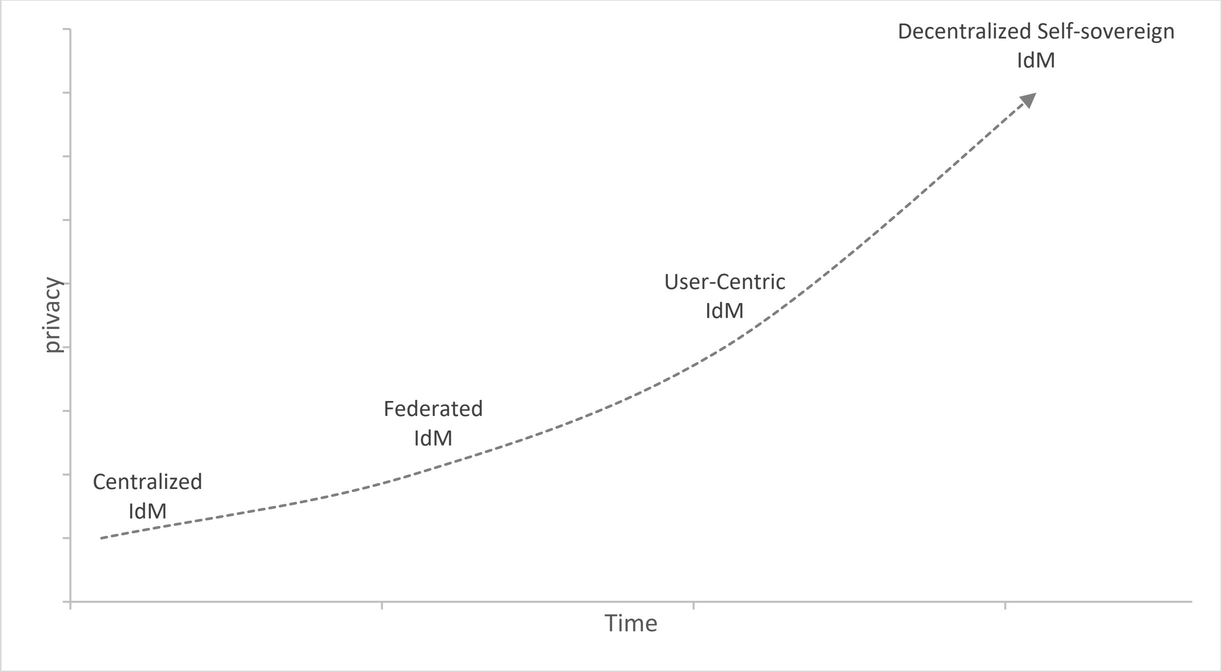


Figure 11. IdM evolution according to privacy preservation capabilities [26]

The traditional central Identity Management (IdM) solutions had many privacy and security challenges. The federated IdM models introduced a partial solution to these challenges by offering Single Sign-On (SSO) and enabling shifting towards a more user-centric approach by leveraging technologies such as OpenID, SAML, and FIDO. [26] The above solution doesn’t fully solve privacy and security issues since the data is still stored on servers and is subject to data leakage and breaches. DSSI can be the ultimate solution as users have full control over their personal identity data without a need for third parties to perform IdM operations. The evolution of IdM methods regarding privacy-preserving capabilities over time is shown in Figure 12.

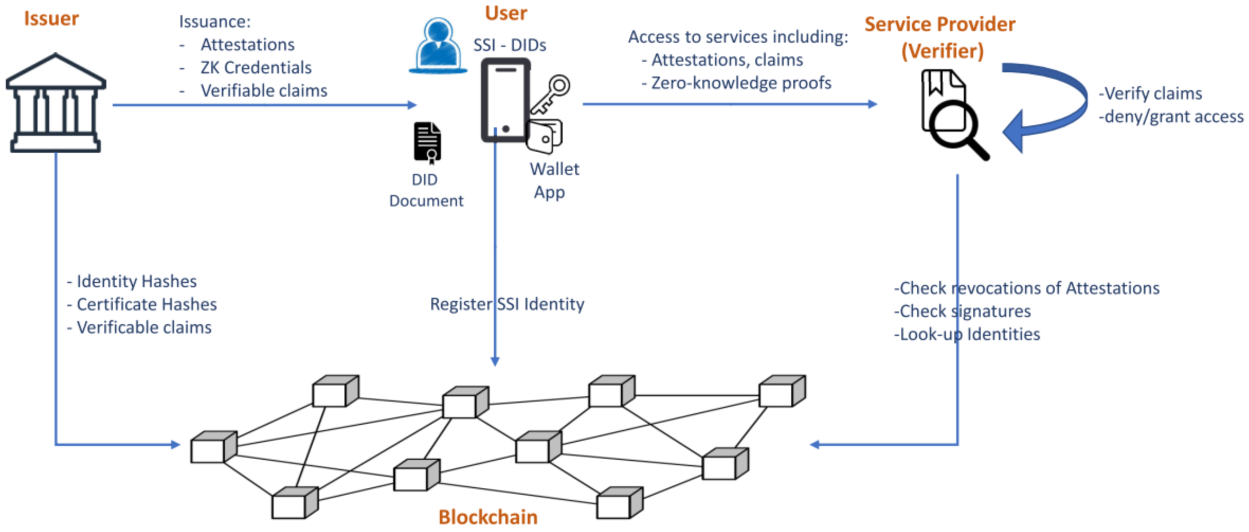


Figure 12. Self-Sovereign Identity Management Model in Blockchain [26]

Bernabe [26] introduces a self-sovereign IdM model in blockchain, as shown in Figure 4. In this model, a User can have DIDs, Verifiable claims, and credentials from the issuer authority, then using ZKPs, a service provider (acting as verifier) can check the attestations and signatures.

Decentralized IdM Projects

Sovrin

An open-source SSI solution based on public and permissioned blockchain technology governed by the non-profit Sovrin Foundation [43], [44]. Sovrin implements Privacy by Design on a global scale, including pairwise pseudonymous identifiers, peer-to-peer private agents, and selective disclosure of personal data using zero-knowledge proof cryptography [45].

uPort

uPort [46], [47] was a Smart contract-based SSI solution founded back in 2015 and now evolved and split into two separate projects named Serto [48], [49] and Veramo [50], [51]. Serto Suite offers decentralized identities and verifiable credentials to businesses, while Veramo is a JavaScript framework geared towards developers that want to use DIDs and verifiable credentials in their applications.

ShoCard

A decentralized trusted identity platform that works across multiple blockchains with three main functions: authentication, authorization, and exchange of attestation. In ShoCard [52], user has control over its own data. User's data can be stored locally, and only the verification of the documents is stored publicly on the blockchain. [53]

Civic

An Identity Verification (IDV) and digital identity platform, also known as Civic Secure Identity Platform (SIP) built on top of Ethereum and Solana blockchains. Civic [54] is mainly focused on addressing privacy and security challenges for attestation of Personal Identifiable Information (PII) in multiple scenarios such as Know Your Customer (KYC) compliance. [55] A full list of supported documents and countries can be found by visiting https://www.civic.com/supported-documents.

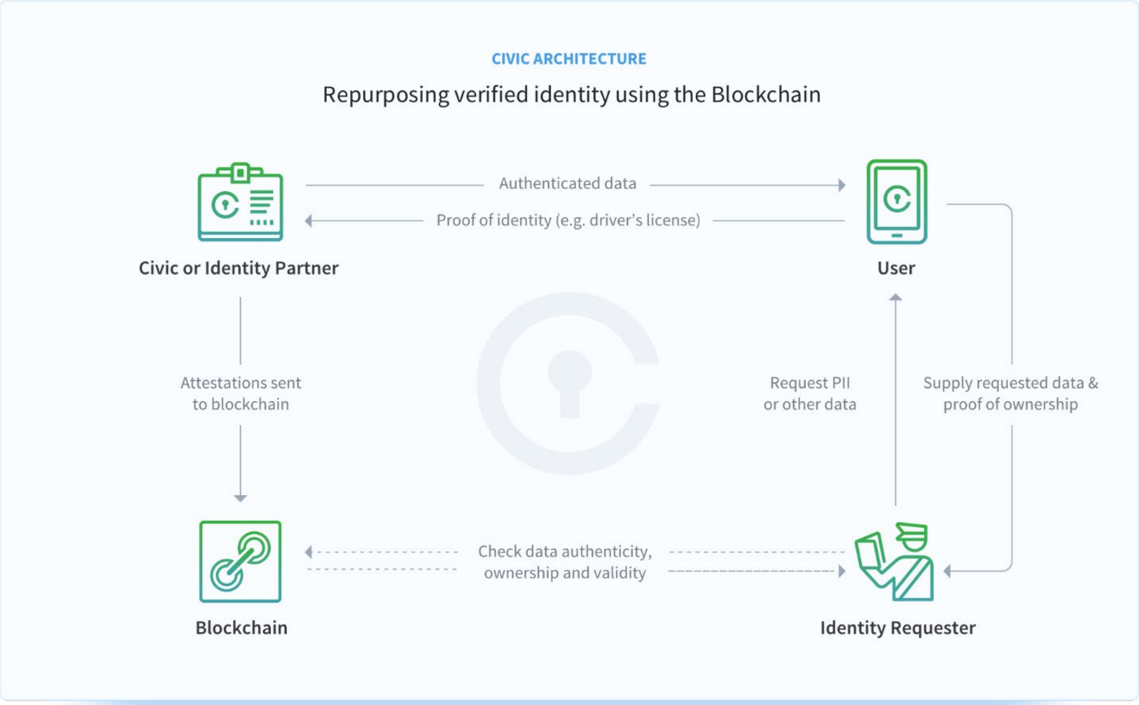

CIVIC ARCHITECTURE
Repurposing verified identity using the Blockchain
Authenticated data
Proof of identity (e.g. driver's license)
Civic or Identity Partner
User
Attestations sent to blockchain
Request PII or other data
Supply requested data & proof of ownership
Check data authenticity, ownership and validity
Blockchain
Identity Requester

Figure 13. Civic Architecture [55]

Secret Network (previously known as Enigma)

The Enigma project [56] was founded back in 2015 by MIT by building new products and systems that would help to speed up the adoption of privacy-first and decentralized technologies. Enigma introduced a peer-to-peer network enabling multiple parties to jointly store and run computations on data while keeping the data completely private. Its computational model was based on a highly optimized version of SMPC, guaranteed by a verifiable secret-sharing scheme [57].

In 2020, The project was rebranded to Secret Network [58]. Secret Network is built with Rust and based on Cosmos SDK, and Tendermint consensus, which means information can be transmitted across a network of blockchains as Cosmos has the inter-blockchain communication (IBC) protocol [59]. Secret Network achieves data privacy through a combination of key management and encryption protocols working within a Trusted Execution Environment (TEE) [60].

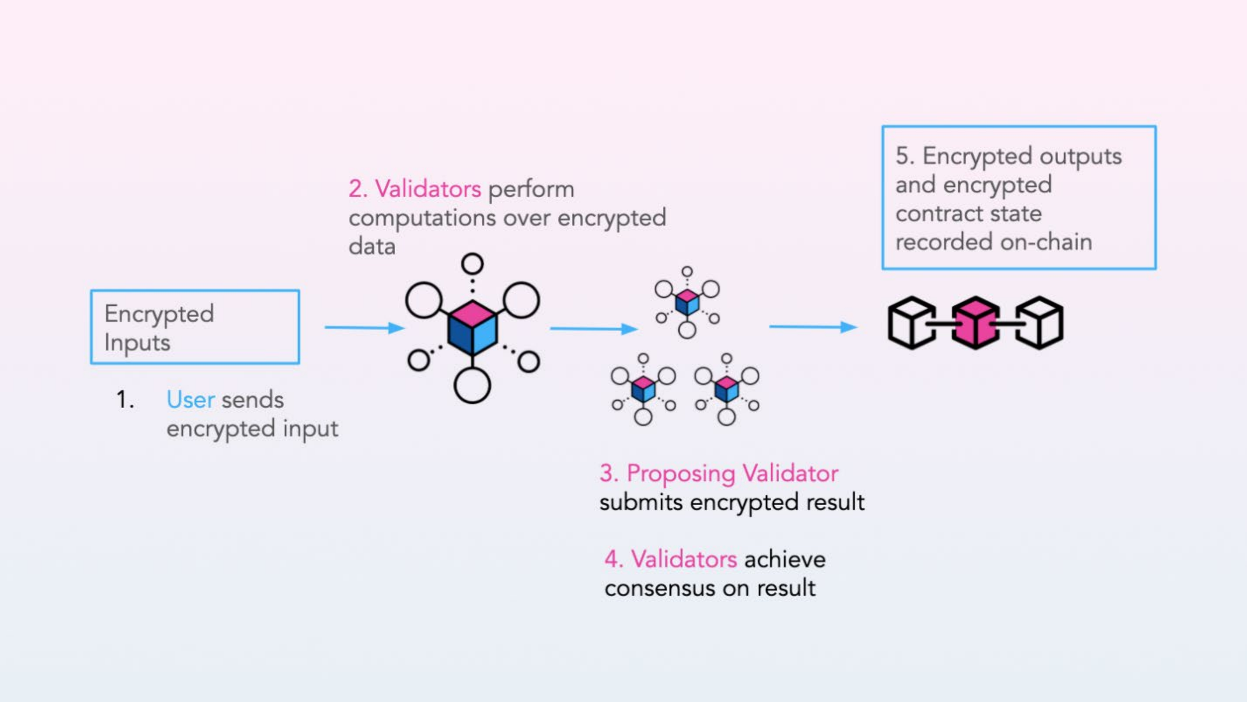

Encrypted Inputs
1. User sends encrypted input
2. Validators perform computations over encrypted data
3. Proposing Validator submits encrypted result
4. Validators achieve consensus on result
5. Encrypted outputs and encrypted contract state recorded on-chain

Figure 14. Secret Network simplified process [60]

The purpose of the new blockchain is to be an open-source protocol enabling a wide range of privacy-preserving tools and applications through programmable privacy, which can ultimately fulfill their initial mission to improve the adoption and usability of privacy-preserving decentralized technologies [61].

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

Conclusions

Blockchain technologies are indeed introducing game-changing and revolutionary changes in many sectors and industries. The decentralized approach to highly centralized services such as identity management could ultimately increase security, privacy, and efficiency. But they are not without flaws; blockchain applications are still young and require a long time to fully explore and comprehend. Meanwhile, contributing to the research and raising awareness would be an essential task for subject experts from many related sectors to analyze and review the technologies from different perspectives

Technologies and techniques like non-interactive zero-knowledge proofs such as zk-STARK enable users to prove their identity and personal information without ever sharing the actual data. Using solutions based on FHE and SMPC, the validated proof or any encrypted message can be processed for data-processing and decision making without decrypting it, for example, Data analysis and data sharing in a regulated industry such as healthcare.

Implementing the privacy-preserving solutions based on blockchain, such as Decentralized IdMs, in the existing ecosystem can introduce the ultimate user-centric approach to achieve higher privacy and security. It also reduces the risk associated with data collection and storage, such as data breaches and identity theft. As the blockchain applications grow over time, more techniques and technologies can be applied to address user and data privacy issues.

Future Work

Reviewing the existing tools proves building a blockchain-based, privacy-preserving solution is not far from achieving. In fact, many of Decentralized IdM platforms seems promising and are already gaining more support. But their approach seems to be focused on building a decentralized solution rather than a privacy-focused one. Developing a globally accepted, privacy-preserving, IdM application based on blockchain technologies only takes the right team and resources to further research and build a solution. Separate future studies in the privacy-preserving application of blockchain technologies can be conducted to explore the possibilities further and address other privacy and security challenges.

The next step to expand on this research would be to acquire research funding and collaborate with computer engineers and scientists to work on developing a blueprint and building the initial model.

REFERENCES

[1] "Consumer Sentinel Network Data Book 2021," Feb. 2022. Accessed: Mar. 28, 2022. [Online]. Available: https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2021

[2] "New Data Shows FTC Received 2.8 Million Fraud Reports from Consumers in 2021," Feb. 2022. Accessed: Mar. 28, 2022. [Online]. Available: https://www.ftc.gov/news-events/news/press-releases/2022/02/new-data-shows-ftc-received-28-million-fraud-reports-consumers-2021-0

[3] Consumer Sentinel Network, "CSN Annual Data Book 2021 Final PDF," 2021. Accessed: Mar. 28, 2022. [Online]. Available: https://www.ftc.gov/system/files/ftc_gov/pdf/CSN%20Annual%20Data%20Book%202021%20Final%20PDF.pdf

[4] B. Auxier, L. Rainie, M. Anderson, A. Perrin, M. Kumar, and E. Turner, "Americans and privacy: Concerned, confused and feeling lack of control over their personal information," 2019, Accessed: Mar. 28, 2022. [Online]. Available: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/11/Pew-Research-Center_PI_2019.11.15_Privacy_FINAL.pdf

[5] ITRC, "Identity Theft Resource Center," 2022. https://www.idtheftcenter.org/ (accessed Apr. 28, 2022).

[6] Identity Theft Resource Center (ITRC), "ITRC Annual Data Breach Report - 2021 in review," 2022.

[7] *Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT ACT) Act of 2001*. H.R. 3162, 2001.

[8] S. Haber and W. S. Stornetta, "Secure names for bit-strings," in *Proceedings of the 4th ACM Conference on Computer and Communications Security*, 1997, pp. 28–35. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/266420.266430

[9] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System.” [Online]. Available: www.bitcoin.org

[10] V. Buterin, “Ethereum Whitepaper,” 2014. Accessed: Dec. 27, 2021. [Online]. Available: https://ethereum.org/en/whitepaper/

[11] A. M. Antonopoulos and G. Dr. Wood, *Mastering Ethereum: Building Smart Contracts and DApps*, First. Sebastopol: O’Reilly Media, Inc, 2018.

[12] G. Iredale, “History Of Blockchain Technology: A Detailed Guide,” Nov. 03, 2020. https://101blockchains.com/history-of-blockchain-timeline (accessed Dec. 27, 2021).

[13] D. Yaga, P. Mell, N. Roby, and K. Scarfone, “Blockchain technology overview,” Gaithersburg, MD, Oct. 2018. doi: 10.6028/NIST.IR.8202.

[14] K. E. Wegrzyn and E. Wang, “Types of Blockchain: Public, Private, or Something in Between,” Aug. 19, 2021. https://www.foley.com/en/insights/publications/2021/08/types-of-blockchain-public-private-between (accessed Jan. 27, 2022).

[15] S. Zhang and J.-H. Lee, “Analysis of the main consensus protocols of blockchain,” *ICT Express*, vol. 6, no. 2, pp. 93–97, Jun. 2020, doi: 10.1016/j.icte.2019.08.001.

[16] Solana, “Solana - Scalable Blockchain Infrastructure: Billions of transactions & counting,” 2022. https://solana.com/ (accessed Mar. 28, 2022).

[17] Ethereum, “Ethereum,” 2022. https://ethereum.org/en/ (accessed Mar. 28, 2022).

[18] Web3 Foundation, “Polkadot - Decentralized Web 3.0 Blockchain Interoperability Platform,” 2022. https://polkadot.network/ (accessed Mar. 28, 2022).

[19] Avalanche, “Avalanche - Blazingly Fast, Low Cost, & Eco-Friendly,” 2022. https://www.avax.network/ (accessed Mar. 28, 2022).

[20] IBM, “What is blockchain security?” https://www.ibm.com/topics/blockchain-security (accessed Jan. 27, 2022).

[21] E. and M. National Academies of Sciences, *Quantum Computing: Progress and Prospects*. Washington, D.C.: National Academies Press, 2019. doi: 10.17226/25196.

[22] D. J. Solove, “Understanding privacy,” 2008, Accessed: Jan. 27, 2022. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1127888

[23] M. Binjubeir, A. A. Ahmed, M. A. bin Ismail, A. S. Sadiq, and M. K. Khan, “Comprehensive survey on big data privacy protection,” *IEEE Access*, vol. 8, pp. 20067–20079, 2019, doi: 10.1109/ACCESS.2019.2962368.

[24] F. A. Viernes, “Stop Saying ‘Data is the New Oil,’” *Medium*, Sep. 14, 2021. https://medium.com/geekculture/stop-saying-data-is-the-new-oil-a2422727218c (accessed Jan. 27, 2022).

[25] The Economist, “The world’s most valuable resource is no longer oil, but data,” *The Economist*, May 06, 2017. Accessed: Jan. 27, 2022. [Online]. Available: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

[26] J. Bernal Bernabe, J. L. Canovas, J. L. Hernandez-Ramos, R. Torres Moreno, and A. Skarmeta, “Privacy-Preserving Solutions for Blockchain: Review and Challenges,” *IEEE Access*, vol. 7, pp. 164908–164940, 2019, doi: 10.1109/ACCESS.2019.2950872.

[27] Y. Lindell, “Secure multiparty computation,” *Commun ACM*, vol. 64, no. 1, pp. 86–96, Jan. 2021, doi: 10.1145/3387108.

[28] S. Goldwasser, S. Micali, and C. Rackoff, “The knowledge complexity of interactive proof systems,” *SIAM Journal on computing*, vol. 18, no. 1, pp. 186–208, 1989, Accessed: Dec. 27, 2021. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.419.8132&rep=rep1&type=pdf

[29] Zerocoin, “Zerocoin Project.” https://zerocoin.org/ (accessed Jan. 27, 2022).

[30] Zcash, “Zcash,” 2021. https://z.cash/ (accessed Jan. 27, 2022).

[31] J. Thaler, “Proofs, arguments, and zero-knowledge.” 2020. Accessed: Dec. 27, 2021. [Online]. Available: https://people.cs.georgetown.edu/jthaler/ProofsArgsAndZK.pdf

[32] Zcash, “Zcash: Parameter Generation,” 2021. https://z.cash/technology/paramgen/ (accessed Dec. 27, 2021).

[33] S. Bowe, A. Gabizon, and M. D. Green, "A multi-party protocol for constructing the public parameters of the Pinocchio zk-SNARK," *Cryptology ePrint Archive*, 2017, Accessed: Dec. 27, 2021. [Online]. Available: https://eprint.iacr.org/2017/602.pdf

[34] E. Ben-Sasson, I. Bentov, Y. Horesh, and M. Riabzev, "Scalable, transparent, and post-quantum secure computational integrity," *Cryptology ePrint Archive*, 2018, Accessed: Dec. 27, 2021. [Online]. Available: https://eprint.iacr.org/2018/046.pdf

[35] R. L. Rivest, L. Adleman, M. L. Dertouzos, and others, "On data banks and privacy homomorphisms," *Foundations of secure computation*, vol. 4, no. 11, pp. 169–180, 1978, Accessed: Dec. 27, 2021. [Online]. Available: https://luca-giuzzi.unibs.it/corsi/Support/papers-cryptography/RAD78.pdf

[36] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proceedings of the forty-first annual ACM symposium on Theory of computing*, 2009, pp. 169–178. Accessed: Dec. 27, 2021. [Online]. Available: https://www.cs.cmu.edu/~odonnell/hits09/gentry-homomorphic-encryption.pdf

[37] IBM, "Homomorphic Encryption Services." https://www.ibm.com/security/services/homomorphic-encryption (accessed Dec. 27, 2021).

[38] R. L. Rivest, A. Shamir, and Y. Tauman, "How to leak a secret," in *International conference on the theory and application of cryptology and information security*, 2001, pp. 552–565. Accessed: Dec. 27, 2021. [Online]. Available: https://people.csail.mit.edu/rivest/pubs/RST01.pdf

[39] D. Chaum and E. van Heyst, "Group signatures," in *Workshop on the Theory and Application of of Cryptographic Techniques*, 1991, pp. 257–265. Accessed: Dec. 27, 2021. [Online]. Available: https://link.springer.com/content/pdf/10.1007/3-540-46416-6_22.pdf

[40] K. M. Alonso and KOE, "Zero to Monero: First Edition," Jun. 2018. Accessed: Dec. 27, 2021. [Online]. Available: https://www.getmonero.org/library/Zero-to-Monero-1-0-0.pdf

[41] KOE, K. M. Alonso, and S. Noether, "Zero to Monero: Second Edition," Apr. 2020. Accessed: Dec. 27, 2021. [Online]. Available: https://www.getmonero.org/library/Zero-to-Monero-2-0-0.pdf

[42] M. Sporny, D. Longley, M. Sabadello, D. Reed, O. Steele, and C. Allen, “Decentralized Identifiers (DIDs) v1.0,” Aug. 2021. Accessed: Dec. 27, 2021. [Online]. Available: https://www.w3.org/TR/did-core/

[43] Sovrin Foundation, “Sovrin,” 2022. https://sovrin.org/ (accessed Jan. 27, 2022).

[44] P. Windley, “How sovrin works,” *Sovrin Foundation*, pp. 1–10, 2016, Accessed: Jan. 27, 2022. [Online]. Available: https://sovrin.org/wp-content/uploads/2018/03/How-Sovrin-Works.pdf

[45] P. Windley, D. Reed, and The Sovrin Foundation, “Sovrin^{TM}: A Protocol and Token for Self-Sovereign Identity and Decentralized Trust,” *Sovrin Foundation - White Paper*, 2018, Accessed: Jan. 27, 2022. [Online]. Available: https://sovrin.org/wp-content/uploads/Sovrin-Protocol-and-Token-White-Paper.pdf

[46] uPort, “uPort.” https://www.uport.me/ (accessed Jan. 27, 2022).

[47] uPort, “Veramo: uPort’s Open Source Evolution,” *Medium*, May 04, 2021. https://medium.com/uport/veramo-uports-open-source-evolution-d85fa463db1f (accessed Jan. 27, 2022).

[48] Serto, “Serto,” 2021. https://www.serto.id/ (accessed Jan. 27, 2022).

[49] Serto, “Serto Suite,” 2022. https://docs.serto.id/docs/intro/ (accessed Jan. 27, 2022).

[50] Veramo, “Veramo - A JavaScript Framework for Verifiable Data | Performant and modular APIs for Verifiable Data and SSI,” 2021. https://veramo.io/ (accessed Jan. 27, 2022).

[51] Veramo, “Veramo Documentation,” 2021. https://veramo.io/docs/basics/introduction/ (accessed Jan. 27, 2022).

[52] Ping Identity, “ShoCard,” 2021. https://www.shocard.com/en.html (accessed Jan. 27, 2022).

[53] D. van Bokkem, R. Hageman, G. Koning, L. Nguyen, and N. Zarin, “Self-sovereign identity solutions: The necessity of blockchain technology,” *arXiv preprint arXiv:1904.12816*, 2019, Accessed: Jan. 27, 2022. [Online]. Available: https://arxiv.org/pdf/1904.12816.pdf

[54] Inc. Civic Technologies, “Civic,” 2022. https://www.civic.com/ (accessed Jan. 27, 2022).

[55] V. Lingham and J. Smith, “Civic Whitepaper,” *CivicTokenSaleWhitePaper.pdf*, 2018, Accessed: Jan. 27, 2022. [Online]. Available: https://tokensale.civic.com/CivicTokenSaleWhitePaper.pdf

[56] Enigma, “Enigma - Securing the Decentralized Web.” https://www.enigma.co/ (accessed Jan. 27, 2022).

[57] G. Zyskind, O. Nathan, and A. Pentland, “Enigma: Decentralized Computation Platform with Guaranteed Privacy,” Jun. 2015, Accessed: Jan. 27, 2022. [Online]. Available: https://arxiv.org/pdf/1506.03471.pdf

[58] Secret Network, “Secret Network - Bringing Privacy to Blockchains, Smart Contracts & Web3.” https://scrt.network/ (accessed Jan. 27, 2022).

[59] Secret Network, “Secret Network Documentation: Protocol, Secret Network and the Cosmos Ecosystem.” https://docs.scrt.network/protocol/protocol.html#secret-contracts-and-use-cases (accessed Jan. 27, 2022).

[60] Secret Network, “Secret Network Documentation: General Overview.” https://docs.scrt.network/overview.html (accessed Jan. 27, 2022).

[61] C. Woetzel, “Secret Network: A Privacy-Preserving Secret Contract & Decentralized Application Platform”, Accessed: Jan. 27, 2022. [Online]. Available: https://scrt.network/graypaper

APPENDICES

Appendix – Glossary of Acronyms and Abbreviations

BaaS	Blockchain as a Service
CSR	Common Reference String
CT	Confidential Transactions (referring to Ring CT)
dApps	Decentralized Applications
DeFi	Decentralized Finance
DH	Diffie-Hellman
DID	Decentralized Identifier
DOB	Date of Birth
DSSI	Decentralized Self-Sovereign Identity (also referred as Decentralized Self-Sovereign Identity Management)
EC	Elliptic Curve
FHE	Fully Homomorphic Encryption
FIDO	Fast ID Online
FTC	Federal Trade Commission
IBC	Inter-blockchain Communication
IdM	Identity Management
IDV	Identity Verification
ITRC	Identity Theft Resource Center
KYC	Know Your Customer
MITM	Man-in-the-middle
MPC	Multi-party Computation
NFT	Non-Fungible Token

NIST	National Institute of Standards and Technology
OSI	Open Systems Interconnection
P2P	Peer-to-Peer
PII	Personally Ideentifiable Information
PoS	Proof of Stake
PoW	Proof of Work
PPDM	Privacy-preserving data mining
RSA	Rivest, Shamir, and Adleman
SAML	Security Assertion Markup Language
SIP	Secure Identity Platform
SMPC	Secure Multi-Party Computation
SSI	Self-Sovereign Identity (also referred as Self-Sovereign Identity Management)
SSN	Social Security Number
SSO	Single Sign-On
TEE	Trusted Execution Environment
UI	User Interface
URI	Uniform Resource Identifier
VM	Virtual Machines
ZKP	Zero-Knowledge Proof
zkSNARK	Zero-Knowledge Succinct Non-Interactive Argument of Knowledge
zkSTARK	Zero-Knowledge Scalable Transparent ARgument of Knowledge