A Survey on Privacy Vulnerabilities in Permissionless Blockchains

Aisha Zahid Junejo[1], Manzoor Ahmed Hashmani[2]
Computer and Information Sciences Department
Universiti Teknologi PETRONAS
Sri Iskandar, Malaysia

Abdullah Abdulrehman Alabdulatif[3]
Department of Computer
College of Sciences and Arts in Al-Rass
Qassim University, Saudi Arabia

***Abstract*—Blockchain decentralization not only ensures transparency of transactions to eliminate need of trusting third party, but also makes the transactions of the network to be publicly accessible to all the participating peers in the network. As a result, data anonymity and confidentiality are compromised making several business enterprises and industrialists hesitant to adopt the technology. Although research community has proposed various privacy-preserving solutions for blockchain, however, they still lack in efficiency resulting in distrust of industries in opting for the technology. This study is conducted for contributing to the existing body of knowledge corresponding to privacy in blockchains. The fundamental goal of this study is to delve into privacy vulnerabilities of the blockchain network in a permissionless setting by identifying non-trivial roots of factors causing privacy breach in blockchain and presenting limitation of existing privacy preserving mechanisms. Studies with superficial comparison of privacy preserving techniques are available in literature but a detailed and in-depth analysis of their limitations and causes of privacy breach in blockchain is yet not done. Therefore, in this paper we first present comprehensive analysis of various privacy breaching factors of the blockchain networks. Next, we discuss existing cryptographic and non-cryptographic solutions in literature. We found out that these existing privacy preserving mechanisms have their own set of limitations and hence are inefficient at current point of time. The existing privacy preserving mechanisms need further consideration of the research community before they're widely adopted and benchmarked. Therefore, in the end, we identified some future directions that need to be addressed to model an efficient privacy preserving mechanism for wider adoption of the blockchain technology.**

***Keywords*—Blockchains; privacy vulnerabilities; cryptographic primitives; anonymity; confidentiality**

I. Introduction

The Blockchain technology is one of the most promising technological trends in the world today. It is a horizontal innovation that has the potential to impact every area of human endeavor [1]. The first application of Blockchains, widely known as Bitcoin, was introduced around a decade ago in October 2008 by S. Nakamoto [2]. Succeeding it, various other cryptocurrencies have been introduced [3] [4] [5] [6]. Initially introduced for the financial transactions of the cryptocurrency, the blockchain technology gradually spread to other sectors as well due to its inherent features. Over the years, the technology has been profusely researched and experimented to bring its benefits to other application areas. The technology has eliminated the need of trusting third parties (i.e., banks) for authorization and record keeping of several transactions by providing transparency [7] and tamper resistance [8]. Transparency in Blockchain networks ensure the availability of the transactions to each node in a distributed network, whereas tamper-resistance makes each recorded transaction to be unmodifiable [9] or removable. Over the years, the technology has been profusely researched and experimented to bring its benefits to other application areas [10]. It is because of the decentralized, immutable and transparent nature of blockchain, that its applications have also been witnessed in non-financial areas like education, internet of things IoT, healthcare, big data, cloud computing, supply chain management, cyber security and so on. The blockchain ledger is written on a base and shared among the participating nodes for verification. This enables even the mutually distrusting nodes verify the data through consensus to achieve consistency and maintain the integrity of the blockchain network. Therefore, despite the fact that blockchain provides greater efficiency, reduced capital costs and greater data protection, it is still vulnerable to privacy issues. The data on the blockchain must be public because different nodes need to calculate and verify the same data so it must be accessible across the network. The transparency and credibility of the data is increased due to public availability of the data, however, it introduces the risk of privacy too as business enterprises and industrial organizations are not willing to make any business details public for adversaries to infer the personal information and extort the clients [11]. It is possible to set access control on the network using permissioned blockchains [12], however, the use of this type of blockchain makes the system more centralized and nullifies the purpose of using decentralized system, altogether. With the recent advancements in blockchain research and the eagerness of industries towards blockchain adoption makes privacy one of the key issues that need to be solved. The research in this paper has been carried out to highlight the issue of privacy in blockchain and the reasons behind it. This will help future researchers to solve the existing issues to get a better privacy protection in blockchain networks for a much wider adoption of this breakthrough technology.

A. Gap Analysis and Contribution

According to the best of our knowledge, various studies [13] [14] [15] have highlighted the importance of privacy preservation in blockchain networks. Although these studies have contrasted existing mechanisms of ensuring privacy, however, they lack comprehensive insight towards possible

factors resulting in privacy disclosure. The study in the paper, therefore, presents comprehensive discussion on root causes of privacy breach in a blockchain network. Based on existing body of knowledge in the domain, we have managed to deduce some meaningful insights that will help research community to design more private blockchain networks. This research study is a multifold: i) describes blockchain technology and its benefits over traditional transaction systems, ii) elaborates the concept and need of privacy in relation to blockchain networks, iii) discusses privacy threats to blockchain and deduces the causes of privacy breach with respect to these threats, iv) discusses existing privacy solution and their limitations, v) suggests future directions to overcome privacy vulnerabilities in blockchain.

B. Organization of the Paper

The organization of the paper is as follows: Section II gives an overview of blockchain and its working mechanism followed by Section III that describes the issue of privacy in various settings of blockchain networks. Section IV discusses various factors causing privacy breach in blockchains. Further, Section V elucidates the existing privacy preserving mechanisms in blockchains and their limitations. Discussion and proposed future directions are presented in Section VI and Section VII concludes the study.

II. Introduction to Blockchain Technology

In 1991, S. Haber and W. S. Stornetta introduced the concept of a cryptographically secured network of blocks [16]. This concept was adopted by Nick Szabo as he worked upon and introduced decentralized digital currency called Bitgold. A decade later, in 2008, the concept was brought into practical implementation by S. Nakomoto [17] in the form of a cryptocurrency that is widely known is Bitcoin. It was since 2008, that the blockchain has been used to implement different cryptocurrencies. Additionally, due to the decentralized, immutable and transparent nature of blockchain, its applications have also been witnessed in non-financial areas like education [18] [19], internet of things IoT [20] [21], healthcare [8] [22] [23], big data, cloud computing, supply chain management [24] [25], cyber security and so on.

Since blockchain networks are distributed, hence the record of transactions is not stored on a single centralized server instead in a case a transaction occurs in the blockchain, it is distributed among all participating nodes where each node maintains a copy of the ledger [26]. This means that there exists thousands and millions of copies of the same blockchain where each node has access to the transaction details. Spreading the information across the network to multiple computers makes the information difficult to be manipulated hence providing transaction record integrity. Fig. 1 depicts the working mechanism of a blockchain network. A user A initiates the transaction that meant for a user B. This transaction is stored on a block and hence the block is created. Once the block has been created it is broadcasted to all participating nodes, also referred as peers, for verification of the transaction. If the transaction is validated by majority of the network, the newly created block is added to the existing chain and a copy of the updated ledger is maintained at each peer for record keeping. This completes a typical blockchain transaction from user A to user B. The authenticity of transactions in a blockchain network is validated via asymmetric cryptography, also widely known as public key cryptography. It is one of the core components of blockchain technology [27]. More information on the types of cryptography can be found in [28] and is not discussed in detail as it is beyond the scope of this paper.

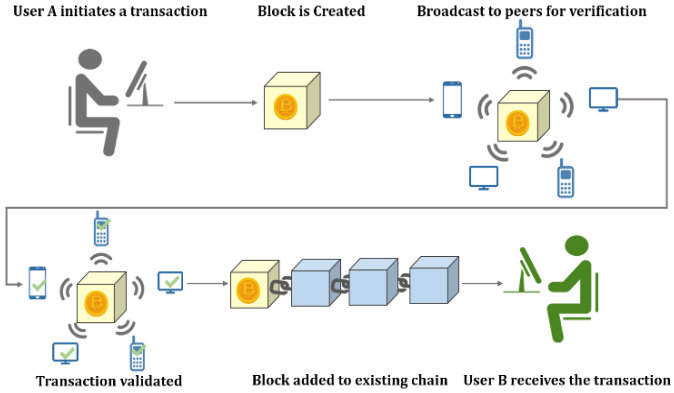

Fig. 1. Blockchain Working Mechanism.

III. Privacy Vulnerability in Blockchain

The blockchain networks are fundamentally transparent and distributed in nature, due to which they are widely being adopted and experimented. However, this means that all the data on a blockchain network is readily available for anyone on the network to view, causing privacy breach.

Blockchain networks can broadly be classified into two categories i.e. permissioned and permissionless blockchains. In a permissionless blockchain, a user requires no permission to enter the network. These kind of blockchains are open for anyone to join and participate. These systems have gained the attention of research community due to their decentralized consensus system [29]. On the other hand, special permissions are required in order to join a permissioned blockchain network. In a permissioned blockchain, the owner has the authority to decide who can join and become a part of the network. This means the blockchain owner has the ability and control to dictate the structure of the network, issue updates of the software, and control whatever operation and process occurs on that blockchain network.

	PERMISSIONLESS	PERMISSIONED
PUBLIC	Anyone	Anyone
	Anyone	Anyone
	Anyone	Authorized User
	Anyone	Authorized User
	Requires Identity Privacy and Data Privacy	Requires Identity Privacy and Data Privacy
PRIVATE	Authorized User	Authorized User
	Authorized User	Authorized User
	Authorized User	Network Operator
	Authorized User	Network Operator
	Requires Data Privacy Only	Requires Data Privacy Only

Join
Read
Write
Commit

Fig. 2. Permissioned vs. Permissionless.

Private and public blockchains can have either permissioned or permissionless setting. This is illustrated in Fig. 2. Public and Permissionless allow anyone to join, read, write and commit to the transactions in the network. This means, all our data, be it personal or not, will be accessible by anyone in the network. This is where the issue of privacy arises. Moreover, in public and permissioned blockchains anyone can join and read the transactions, however only authorized users can write or commit. This improves trust in the blockchain but still doesn't guarantee the privacy of our assets. Similar is the case in Private and Permissionless blockchains. Lastly, in private and permissioned blockchains, although all users are known to the authorities, but this still doesn't guarantee the privacy of the data being transacted. So whatever type of blockchain it is, it does require privacy guarantee.

IV. Causes of Privacy Breach in Blockchain

Blockchains provide efficiency, reduced costs, transparency and trust but is still prone to privacy breach. For wider adoption, the privacy of blockchain networks must be strengthened. This section covers several causes resulting in privacy disclosure in blockchain networks.

A. Anonymization Inefficiency

In blockchain networks, anonymization refers to hiding the identity of the user. Anonymity is achieved when:

- Public address of the user cannot be mapped to his real identity.
- Blockchain transactions do not contain any personal identifiable information (PII).

Despite of blockchain claims of anonymity, it does not provide enough privacy. Several techniques are available in literature through which the anonymity of a blockchain network can be broken to identify the actual participants involved in a certain transaction. The phenomenon of disclosing user anonymity is known as deanonymization. In deanonymization, analysis of the network and network listening can help identify the blockchain user by unmasking him [13]. Further elaboration on deanonymizing blockchain users is presented in following subsections. Note that since cryptocurrencies are the first and widest applications of blockchain networks, hence the discussion carried out in following few sections will mainly focus cryptocurrencies to understand privacy mechanism and vulnerable areas of the technology. The same idea can further be applied to different applications.

1) Deanonymizing via network analysis: Each successful transaction in blockchain is added to transaction network where every node represents a transaction, and every (directed) edge represents a flow of data from an output of one transaction to an input of another. Analyzing the network relationships can be used to deanonymize a user's identity, thereby compromising the privacy. Since blockchain is a P2P network, hence IP address of nodes can be leaked [13] while transaction broadcasting.

2) Deanonymizing via address clustering: It is possible for transaction contents, transactions relationship with other transactions and the way transaction is broadcasted, to unintentionally leak information about the parties involved in the transaction to interested third parties. It is in fact noticed that various interested third parties systematically gather this kind of information to analyze various user patterns for multiple reasons including market research, competitor analysis, compliance and law enforcement. This analysis can (though not easily) be carried out using address clustering. The idea is to partition the set of addresses involved in a transaction to as many numbers of subsets as possible. Each subset, known as address cluster, most likely corresponds to the same entity. By combining address clusters with address tagging and graph analysis [30], the activity in blockchain can be effectively analyzed.

3) Deanonymizing via transaction fingerprinting: Another threat to anonymity is transaction fingerprinting. Androulaki investigated Bitcoin privacy provisions in a university setting. A simulator to mimic Bitcoin system was used and the results depicted that about 40% of the users' identities can be recovered despite of using Bitcoin's privacy measures [31].

Table I shows various deanonymization attacks on blockchain based cryptocurrencies.

TABLE I. Deanonymization Attacks on Cryptocurrencies

S.No	Paper Title	Privacy Threat	Success Rate	Test Case
1	An analysis of anonymity in Bitcoin using P2P network traffic [33]	Network Analysis	>90%	Bitcoin
2	Deanonymization of clients in Bitcoin P2P network [34]	Network Analysis	11% - 60%	Bitcoin,
3	Deanonymization and linkability of cryptocurrency transactions based on network analysis [35]	Network Analysis		Bitcoin, Zcash, Dash, Monero
4	Data-Driven De-Anonymization in Bitcoin [36]	Address Clustering	68.59%	Bitcoin
5	Evaluating User Privacy in Bitcoin [31]	Transaction Fingerprinting	40%	Bitcoin

B. Transaction Pattern Linkability

Transaction information following through the public network can be used to reach out to statistical distributions on Cryptocurrencies revealing some new regulation within blockchain applications [13].

1) Threat of transaction graph analysis: M. Moser et al. [32] developed a framework based on transaction graph analysis to deanonymize the identities of users from publicly available transaction information in Bitcoin. Monero was taken as test case in the study and was empirically evaluated. Mix-ins used in Monero resulted in about 62% of the

transactions being unshielded to chain reaction i.e. deducing the actual input by elimination method. Moreover, The sampling of mix-ins in Monero is done in such a way that it gets easier to distinguish them from the real coins using their age distribution; in short, the real input is usually the "newest" input.

The authors estimated this phenomenon to guess the real input with around 80% accuracy. Further, each transaction in cryptocurrencies have some number of inputs and outputs that consume and create new coins respectively to conserve the total balance. Each input spends the new coins created in prior transaction and hence a transaction graph is formed. The public nature of blockchain data poses a potential privacy hazard to users. Since each transaction is publicly broadcast and widely replicated, any potentially identifying information can be determined for even years after a transaction is committed. The study depicted that a huge amount of data in Monero is traceable.

In another study [37] the authors focused on the typical behavior of users, the way they acquire spend their bitcoins, the balance of bitcoins they keep in their accounts, the way they move bitcoins between their various accounts in order to better protect their privacy. In addition, the research study isolated all the large transactions in the system, and discovered close relation of all these transactions to a single large transaction that took place in November 2010, even though the associated users apparently tried to hide this fact with many strange looking long chains and fork-merge structures in the transaction graph. Similarly, another study was carried out to test transaction linkability with the test case being Monero, again. In this study, three attack routines were developed to test against Monero's privacy guarantee. The results of the study depicted in 88% of the cases it was easy to determine the origin of funds transferred.

2) Web payment: When a user makes a payment through web or online wallets, the consumer identity is prone to be linked to his real identity via browser cookies. When the user pays with a cryptocurrency, the service provider can link the real identity to the token history in the blockchain which also states that the attack is resilient against mixing mechanisms like CoinJoin [14].

In [38], two attacks are presented. The first attack shows that web trackers can extract substantial amount of information for advertising and analytics purposes when the user makes purchases on shopping websites. This information is enough to identify the blockchain transaction uniquely for linking it with the web cookies of the user to further reveal user's identity. The second attack depicts that by linking even two purchases of the same user, the web tracker can identify his cluster of addresses even if anonymity techniques of blockchain such as CoinJoin are deployed. Moreover, it is possible to apply the attacks to past purchases as well. Thus, in the study, it is shown that third party web trackers have the ability of deanonymizing the cryptocurrency users.

A summary of studies carried out under this kind of privacy threats is given in Table II.

C. Crisis of Private Key Theft

Private keys in a blockchain network are very critical to ensure the security and privacy of the user because these keys are used for signing each transaction in the network. Participant's assets are controlled through private key in the blockchain systems. Hence, it is very important that proper key management systems [39] are enforced. If compromised, it can not only lead to privacy leakage but may also result in identity theft.

Although, private key allows a user to have sovereignty over his assets, however it comes under the responsibility of securing and managing one's own private keys. Currently, there are no efficient mechanisms for recovery of the keys in a case of loss. Table III summarizes some of private key theft incidences compromising the security and privacy in blockchain systems.

TABLE II. TRANSACTION PATTERN LINKABILITY ATTACKS ON CRYPTOCURRENCIES

S.No	Paper Title	Privacy Threat	Success Rate	Test Case
1	A Traceability Analysis of Monero's Blockchain [40]	Transaction Graph Analysis	88%	Monero, RingCT
2	Quantitative Analysis of the Full Bitcoin Transaction Graph [37]	Transaction Graph Analysis	62%	Monero
3	When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies [38]	Web Payment	Attack I: 90% Attack II: T = 2 – 89% T = 3 – 99%	Bitcoin

TABLE III. PRIVATE KEY THEFT INCIDENTS

S.No	Incident	Amount	Year	Victim
1	A hacker took possession of the administrative account was hacked and private keys were stolen. BTC price was changed to 1 cent and bought BTC from Mt. Gox users.	2643 BTC	2011	Mt. Gox
2	Attacker got access to bitcoinica database, obtained private information of users for theft.	38,000 BTC	2012	Bitocoinica
3	Unencrypted Private Keys stored online for backup were stolen	24,000 BTC	2012	Bitfloor
4	The attacker under the nickname Lucky7Coin inserted the Trojan code into the code of Cryptsy—a cryptocurrency exchange. A hacker got access to BTC and LTC keys.	13,000 BTC 300,000 LTC.	2014	Cryptsy
5	Hackers infected the internal network of the exchange with a virus that was transmitted through email, and it allowed them to steal private keys.	523M NEM	2018	Coincheck
6	Phishing and malware tactics were used to steal user 2FA codes and API keys alongwith customers' private details.	7,000 BTC	2019	Binance

V. Inefficiency of Existing Privacy-Preserving Frameworks

Blockchain technology has two categories when it comes to preserving privacy. The first category involves protecting the identity of the user by assigning him complete anonymity while making transactions. The second category involves protecting the transaction data from unauthorized entities and hackers thus maintaining data confidentiality. The classification of various privacy preserving techniques surveyed in the literatures are depicted in Fig. 3 and detailed in the subsequent section. The classification is done based on which technique contributes towards achieving what kind of privacy in blockchains.

The privacy preserving frameworks, reviewed in literature can broadly be classified into two categories, i.e.:

- Mixing Methods: Mixing methods or services are used to retain the transaction data privacy of the blockchain networks.
- Cryptographic Primitives: Cryptographic primitives are mathematical functions that are used in cryptography to verify data authenticity.

A. Privacy Vulnerability in Mixing Services

Link between sender and receiver in a blockchain network can be known by analyzing the publicly available content. Introduction to mixers provides a solution to the stated problem. The concept of mixing service was first presented in [41] by Chaum. It allows users to hide who a participant communicates with as well as the content of the communication.

In Fig. 4, the basic architecture of a mixer is depicted. There are two types of mixing services, i.e., centralized mixing and decentralized mixing. Both concepts are elaborated:

1) Centralized mixing: Multiple mixing websites are available for use. These offer mixing of the transactions anonymously on exchange of mixing fees. The websites swap the transactions among various users so that the relationship between incoming and outgoing transactions can be hidden. Centralized mixing suffers from various limitations (discussed in section 3.3) including the mixing server being prone to denial of service (DOS) attacks as the server remains a single point of failure. Resultantly, it becomes an obstruction of the distributed blockchain network

2) Decentralized mixing: Decentralized mixing overcomes the limitations of centralized mixing which makes it vulnerable to DOS attack. A decentralized mixing pattern is proposed to enable a set of mutually untrusted peers to publish their messages simultaneously and anonymously without the need of a third-party anonymity proxy. Moreover, decentralized mixing eliminates the need of paying mixing fees. CoinJoin [42] and MultiParty [43] Computation are only two methods in literature that has successfully implemented decentralized mixing services.

3) Critical analysis of mixing services: Although mixing services can provide a substantial amount of identity privacy, however, it has its own set of concerns which shall be taken into account before opting out for such a privacy preserving mechanism. These issues are discussed below:

a) Waiting delay: In order to use mixing services, user must wait for other participants to swap their transactions in order to hide and relationships between a transaction inputs and outputs. This incurs high waiting delay for a transaction to be completed.

b) Third party involvement: Since mixing servers are usually websites or other third-party software, hence they're not an appropriate solution to the privacy vulnerability of blockchain networks.

c) Malicious mixing services: Although mixing services hide the relationship between a user's transaction's input and output from an adversary, however, the server itself knows about all the input-output pairs and hence, the privacy in this scenario solely relies on how honest the intermediary is and becomes prone to breaches.

d) Mixing fees: Mixing services usually incur cost of hiding the identities of the users via mixing.

B. Privacy Vulnerability in Cryptographic Primitives

There are two categories of cryptographic algorithms when it comes to blockchain networks. The first ones are primary, which are important for data transaction and communication in blockchain networks, the second ones are optional which are used for preserving and enhancing user and transaction data privacy [44] in blockchain networks.

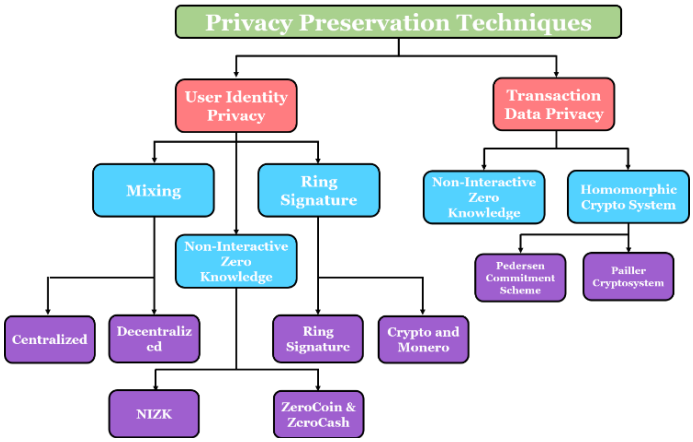

Fig. 3. Classification of Privacy Preserving Techniques.

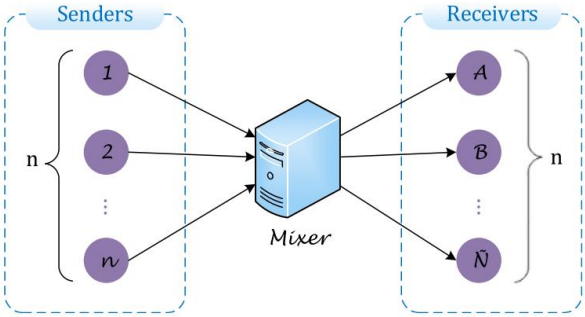

Fig. 4. Mixing Service Architecture [57].

In permissionless blockchain networks, any peer is able to join the network as participant at any point in time. No centralized authority manages or supervises that who joins the network or who should be banned from the network in permissionless scenario. This results in the content of the blockchain to be readable by any peer in the network. However, using optional cryptographic primitives, a permissionless blockchain network can be designed in such a way that privacy of the network in enhanced and each peer gets only relevant information [44]. Currently, the most widely used technologies to achieve blockchain privacy are ring signatures and zero-knowledge proofs.

1) Ring signature: In cryptography various kinds signatures, such as blind signature, ring signature, group signature and DC-nets, from which only ring signature and its variants are used to achieve anonymity in blockchains [44].

Ring signature was introduced in 2001 by Rivest et al. [45]. The concept behind ring signature is that a user chooses a set of participants to create a ring, including himself. Each participant in the ring has a public key. The user initiating the ring signs the message with his/her private key and public keys of all participants. Verifying node knows that one of the members signed the message but can't tell who actually signed it. Hence, anonymity is achieved.

The working mechanism of ring signature is illustrated in Fig. 5. The signature is analogous to the signature for a cheque in joint bank account where all participants sign the transaction with their public keys along with the originator's private key. After each participant of the ring has signed the transaction, it goes further for validation and verification.

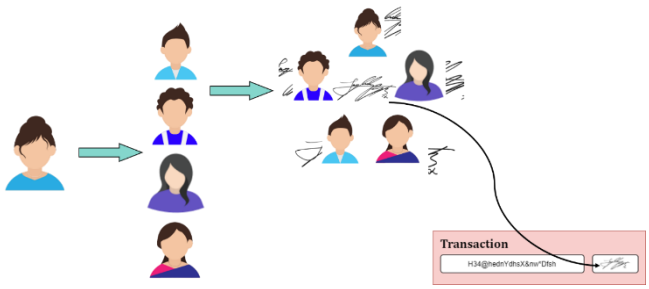

Fig. 5. Transaction Signing in Ring Signature.

Two basic advantages provided by ring signatures include unforgeability and anonymity [46]. Anonymity can further be sub divided into two properties i.e. unlikability and untraceability [44]. Unlinkability refers to the verifier not being able to decide the link between two transactions whereas untraceability refers to the signer not being identified. These properties have led to development of several ring based privacy preserving protocols [47] [48] [49] [50] [51] which are widely used blockchain networks.

A signature scheme known as linkable spontaneous anonymous group (LSAG) was proposed in 2004 [47]. It is a variant of linkable ring signature in which groups are formed spontaneously without any group manager. The concept of ring signature was extended in [48] into traceable ring signature where an issue related tag was added to the signature. This idea was further adopted in [49] for the design on Ring-Coin with improved efficiency. In this case, anyone in the ring, pretending to be another person to sign the same message, would face the risk of revealing his/her identity immediately. This idea was further adopted for preventing double-spending attack in blockchain and became the basis of CryptoNote [50] with a slight modification.

Furthermore, a concept of confidential transaction, using homomorphic commitment protocol, was proposed [51] for hiding transaction amounts. Later, three techniques i.e. ring signature, confidential transaction and multilayered linkable spontaneous anonymous group signature (MLSAG) [52] were combined to form Ring Confidential Transactions (RingCT), with its implementation being in Monero. Besides these, one-time signature [53], borrowmean signature and multisignatures are also used for preserving privacy in blockchain networks [44].

2) Critical analysis of ring signature: Monero [50], based on ring signature is considered to be the most efficiently privacy preserving cryptocurrency, however, Monero (due to vulnerabilities in the architecture of ring signature) also faces privacy issues. Some issues with ring signature include:

a) Large ring size: The size of the ring is directly proportional to the number of participants involved in the ring; this increases the ring size. To keep the ring size limited, usually the no. of participants that can take a part in ring formation is limited. This reduces the anonymity set size, hence increasing the risk of deanonymization.

b) Lack of scalability: Transaction size in ring signature is large – almost thousands of bytes per transaction. This will require more storage space to keep the records of the entire blockchain, hence compromising the scalability of blockchains.

c) Transaction timing attack: When a user creates the ring for his transaction, he usually collects other transactions of the same denomination available in the blockchain. Since each transaction in blockchain is time stamped, hence the newest created transaction in the anonymity set is considered to be the one to be redeemed. A study [40] depicts that 98% of the transactions are prone to time attack for traceability.

3) Zero-knowledge proof: Zero-knowledge protocols, introduced in 1980s [54], are one of the most widely used cryptographic techniques to enable the transfer of assets across a distributed, peer-to-peer blockchain network with improved privacy. The goal of zero-knowledge proofs is to prove the validity of a transaction with zero knowledge provided to the verifier about the transaction. The concept involves the certifier to formulate a formal proof to prove that a certain assertion is true without the need of providing any additional and useful information to the verifier [15]. A variant of ZKP, known as Non-Interactive Zero-knowledge Proof (NIZK proof), is widely used in blockchains as it eliminates the need of to and fro communication between the prover and the verifier and instead, requires only one time message to be sent from prover to the verifier. It is important to remark that not all ZKP schemes are non-interactive. Most of the ZKP protocols available in literature are interactive. Usually, in

ZKP scenario, the prover is required to answer various challenges sent by verifier, resulting in multiple rounds of communication. However, for blockchains and other distributed ledger technologies (DLT), it is desirable to avoid the communication because either (i) validating nodes can't properly agree on how to choose those challenges, since in many constructions we have to choose them randomly, while the verification algorithm must be deterministic in order to reach consensus; or (ii) because it would make the communication complexity of the system very poor. This property makes it suitable for anonymous and distributed verification of messages in blockchains.

The concept first appeared in [54] and is accepted for creating privacy preserving protocols in blockchain networks. NIZK proofs must meet the following three properties:

- Completeness: Everything that is true has a proof.
- Soundness: Everything that can be proved is true.
- Zero knowledge: Only the proven statement is revealed.

Zerocoin, introduced in [55] uses NIZK proof cryptography for providing anonymity by preventing transaction graph analysis i.e. by breaking the trace of coins. However, it fails to provide complete anonymity due to following reasons:

- Fixed denomination coins are used.
- Before payment is made, anonymous coins need to be converted into non-anonymous ones.
- The amount of transactions, or other metadata is not hidden.

To overcome the limitations of Zerocoin, zerocash was introduced [56]. Identity and transaction privacy were simultaneously provided in Zerocash to overcome the limitations of Zerocoin. It uses anonymous coins to provide privacy in blockchains. Further, size of transaction and time of verification of transactions were also significantly reduced. Zerocash uses ZK-SNARKS. However, the NIZK protocol incurs high computation overheads, especially in the proof generation phase of zk-SNARKs protocol used in Zcash.

4) Critical analysis of zero knowledge proof: Despite of providing both identity privacy and data privacy, ZKPs still have not perfected at preserving privacy in blockchain networks. A few issues with ZKPs include:

a) Trusted Setup Problem: The working of ZKPs involve a parameter generator that can issue prover and verifier keys to verify a transaction. This is where vulnerability to privacy breach arises as it is very significant to consider who to trust for parameter generation and how to ensure no record keeping at the generator. If compromised, this may result in forgery of the data.

b) High Computation Overhead: Theoretically, ZKPs achieve the highest level of anonymity and transaction privacy protection for the blockchain but at the expense of high computational costs it requires when it generates the transaction proofs.

c) Prone to deanonymization: A study [57] empirically shows that 98% transactions in Zcash are linkable.

VI. Discussion for Way Forward

Maintaining privacy in blockchain based networks is very significant for its wide acceptance and adoption as shown in the literature. Besides the actual data, metadata also flows through the blockchain network. This metadata can be used to infer additional information about the users participating in the transaction. Additional information inferred may include the identity of the user and this identity unmasking can further reveal all the transactions related to the user. In other words, even with the most powerful privacy preserving mechanisms, this metadata still flows through the network. This is one of the biggest challenges for any privacy protecting approach that might be used in public permissionless blockchain networks. Adding mix-ins to transactions do not have an impactful effect either. Temporal analysis makes it evident that timing plays a major role in analysis of user identity thereby nullifying the effect of mix-ins. Analyzing transaction relationships, patterns, time and links, it becomes easier to trackback the headnode and determine the identity of a person. Once the identity of an individual is leaked, all the corresponding transaction information of the individual also gets prone to leakage.

In certain organizations, it is not desirable to make the confidential data publicly available, for instance patient records in healthcare, sensor data in IoT devices, private goods' information on supply chain management systems, business transactions in financial sector and so on. Hence, keeping privacy intact when blockchains are deployed for those applications, has a great significance. If privacy is not ensured, the integration of blockchain in such application areas may not progress and soon come to a halt. Setting access control is possible by permissioned blockchain, however, using those kinds of blockchains nullifies the purpose of using a decentralized system altogether.

Privacy in a blockchain network can be preserved in various ways but the most prominent one in literature is preserving privacy through the use of efficient cryptographic primitives. A brief summary of type of privacy offered and limitations of existing privacy protecting mechanisms' implementations is presented in Table IV.

It can clearly be seen from the table that existing approaches have a number of limitations and thus need further research for reduction of the privacy risk in blockchain systems. Hence, a few research directions are presented that can be investigated further.

A. Transparency vs. Privacy

Blockchain is transparent by virtue of its design. Transparency, however, can be a double-edged sword when it comes to blockchain transactions. On one hand, blockchain is trusted for its transparency whereas on the other hand, this results in serious privacy concerns for a variety of potential application domains. The desire of stronger privacy in some applications leads to limited usage of the technology. Hence, the biggest challenge to achieve privacy in blockchain systems is finding the correct balance between the degree of transparency and the degree of privacy leveraged.

TABLE IV. SUMMARY OF EXISTING PRIVACY PRESERVATION TECHNIQUES

S.no	Privacy Preservation Method	Type of Privacy		Fundamental Framework	Limitations
		Identity	Data		
1	Mixing websites, (Cryptomix, Bitmix.Biz, SmartMix)	✓	×	Centralized Mixing	Long waiting delay High Mixing Fees (4-5% of the transaction for these particular types) Prone to DDoS and Sybil attacks
2	Centralized Tumblers [58]	✓	×	Centralized Mixing	Long waiting delay High Mixing Fees Prone to money laundry attacks Depends on the trusted party Cannot guarantee safety from theft
3	CoinSwap [59]	✓	×	Centralized Mixing	Long waiting delay High Mixing Fees Prone to DDoS and Sybil attacks No proof that the mixer is not storing transaction record
4	CoinJoin [42]	✓	×	Decentralized Mixing	Long waiting delay Prone to DDoS and Sybil attacks Lacks internal unlinkability
5	CoinShuffle [60]	✓	×	Decentralized Mixing	High communication and computation overhead Can be frustrated by dishonest participants Prone to Sybil attack
6	CoinParty [43]	✓	×	Decentralized Mixing	2/3 users are honest (in theory) Lesser theft prevention
7	RingCT [53]	✓	×	Ring Signature	Large Transaction Size Increasing no. of participants increases ring size
8	CryptoNote [50]	✓	×	Ring Signature	Limited Ring Size Lacks Scalability due to larger transaction size Smaller anonymity set
9	Zerocoin [55]	✓	✓	Zero-knowledge Proofs	Requires larger proof size (Computationally complex) Leakage of trusted setup parameters can lead to forgery of coins Requires fix denominations Requires trusted setup
10	Zerocash [56]	✓	✓	Zero-knowledge Proofs	Computationally intensive Leakage of trusted setup parameters can lead to forgery of coins Requires trusted setup

B. Scalability

Some privacy preserving techniques provide a sufficient amount of privacy for a wide variety of applications. In addition, advanced versions of some of the existing techniques i.e. ring signatures and zero-knowledge proofs (ZKP) provide both user privacy and data content privacy. However, this privacy protection is done at the cost of scalability of the network. Scalability, itself, is one of the major concerns in the technology of blockchain these days, hence, industrialists do not opt for the privacy solutions that further increase the issue. The need of scalable solutions make it another significant challenge in terms of privacy protection of user and user assets. Therefore, researchers should delve further into the cryptography of these techniques to find out the loopholes in existing techniques. The identified loopholes will further help the researchers to model scalable privacy preserving mechanisms.

C. Private Key Management Systems

Loss or theft of private is another major issue that may result in privacy breach of the user and loss of user assets associated with the key. Proper private key management systems should, therefore, be incorporated. Moreover, mechanisms to recover or report the lost keys should be brought into practical implementation.

VII. CONCLUSION

Invention of blockchain eliminated the need of trusting a third party for record keeping and transaction verification. Blockchains promote transparency by introducing publicly verifiable transactions. However, this transparency has led the blockchain community to an emerging issue of privacy. Privacy in blockchain refers to safeguarding the identity of the user involved in a transaction and protecting the secrecy of transaction data. Although researchers and industrialists have proposed some privacy preserving mechanisms over the years, however, these mechanisms are still prone to privacy breaches and do not provide complete privacy. For instance, mixing services and ring signatures can provide user identity privacy only and does not provide transaction data privacy. Similarly, homomorphic cryptosystems aim at providing transaction data privacy but does not provide user identity privacy. Moreover, although ZKPs provide both kinds of privacy in blockchains but it does so at the cost of system performance. Poor

performance of the techniques restricts universal adoption of blockchain technology. Hence, the need for a more efficient privacy preservation framework that doesn't only retain user identity and transaction data privacy, but also ensures the performance of the system doesn't lag arises. For development of an effective solution to problem of privacy in blockchain, understanding the root cause of the issue is important. Therefore, in this study we have highlighted some privacy breaching causes by the virtue of blockchain design. These causes include (i) additional information flowing through the network that aids in deanonymizing a blockchain user; (ii) linking the time and pattern of transactions; and (iii) absence of effective private key management systems in the case of private key thefts. In order to be completely benefitted by the variety of features that blockchain has to offer, it is essential that the privacy in blockchain systems shall be strengthened.

ACKNOWLEDGMENTS

The authors would like to extend their gratitude to Universiti Teknologi PETRONAS for provision of necessary equipment and resources to carry out the research.

REFERENCES

[1] Junejo A.Z., Memon M.M., Junejo M.A., Talpur S., Memon R.M. (2020) Blockchains Technology Analysis: Applications, Current Trends and Future Directions—An Overview. In: Peng SL., Son L., Suseendran G., Balaganesh D. (eds) Intelligent Computing and Innovation on Data Science. Lecture Notes in Networks and Systems, vol 118. Springer, Singapore. http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-981-15-3284-9_47.

[2] Satoshi Nakamoto, "Bitcoin: A peer-to-Peer Electronic Cash," 2009.

[3] C. Lee, "Litecoin," 2011. Whitepaper.

[4] "Peercoin–secure & sustainable cryptocoin," August 2012. Available: https://peercoin. net/whitepaper.

[5] S. King, "Primecoin," 7 July 2013.

[6] I. Grigg, "EOS: An Introduction,"

[7] Guy Zysking, Oz Nathan, Alex 'Sandy' Pentland, "Decentralizing Privacy: Using Blockchain to Protect Personal Data.," in Security and Privacy Workshops, San Jose, 2015.

[8] Asad Ali Siyal, Aisha Zahid Junejo, Muhammad Zawish, Kainat Ahmed, Aiman Khalil, Gerogia Sorsou, "Applications of Blockchain Technology in Medicine and Healthcare: Challenges and Future Perspectives.," MDPI Cryptography, vol. 3, no. 3, 2019.

[9] D. Yaga, Peter Mell, N. Roby, K Scarfone, "Blockchain Technology Overview," NIST, 2018.

[10] V. K. Supriya Thakur, "Blockchain and Its Applications – A Detailed Survey," International Journal of Computer Applications, vol. 180, no. 3, 2017.

[11] Yuchong Cui, Bing Pan, Yanbin Sun, "A Survey of Privacy-Preserving Techniques for Blockchain," in Artificial Intelligence and Security, New York, USA, 2019.

[12] T. K. Sharma, "Permissioned And Permissionless Blockchains: A Comprehensive Guide," Blockchain Council, 13 November 2019. [Online]. Available: https://www.blockchain-council.org/blockchain/permissioned-and-permissionless-blockchains-a-comprehensive-guide/. [Accessed 23 March 2020].

[13] Qi Feng, Debiao He, Sherali Zeadally, Muhammad Khurram Khan, Neeraj Kumar, "A survey on privacy protection in blockchain system," Journal of Network and Computer Applications, vol. 126, pp. 45-58, 2019.

[14] Jorge Bernal Bernabe ; Jose Luis Canovas ; Jose L. Hernandez-Ramos ; Rafael Torres Moreno ; Antonio Skarmeta , "Privacy-Preserving Solutions for Blockchain:," IEEE Access, vol. 7, pp. 164908 - 164940, 2019.

[15] R. Zhang, R. Xue, and L. Liu, ""Security and privacy on blockchain,," CoRR (to appear in ACM Computing Survey), vol. abs/1903.07602, 2019.

[16] H. S. a. S. W., "How to Time-Stamp a Digital Document.," Journal of Cryptology , pp. 99-112, 1991.

[17] G. Wood, Ethereum: A secure decentralised generalised transaction ledger, 2014.

[18] J. D. M Sharples, "The Blockchain and Kudos: A Distributed System for Educational Record, Reputation and Reward.," in Adaptive and adaptable learning, 2016.

[19] D. Skiba, "The potential of Blockchain in education and health care.," Nursing Education Perspectives, vol. 38, no. 4, pp. 220-221, 2017.

[20] M. S. Ali, M. Vecchio, M. Pincheira, K. Dolui, F. Antonelli, and M. H. Rehmani, "Applications of blockchains in the Internet of Things: A comprehensive survey," IEEE Communication Surveys and Tutorials, vol. 21, no. 2, pp. 1676-1717, 2019.

[21] M. A. P. I. 2.-0. O. I. 2.-0. p. 1.-6. V. M. H. Miraz, "Applications of Blockchain Technology beyond cryptocurrency," Annals of Emerging Technologies in Computing (AETiC),, vol. 2, pp. 1 - 6, 2018.

[22] C. Esposito, A. De Santis, G. Tortora, H. Chang and K. R. Choo, "Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy," IEEE Cloud Computing, vol. 5, no. 1, pp. 31-37, 2018.

[23] Ekblaw, A.; Azaria, A.; Halamka, J.D.; Lippman, A., "A Case Study for Blockchain in Healthcare: "MedRec" prototype for electronic health records and medical research data," in IEEE Open and Big Data Conference, Vienna, 2016.

[24] S. &. M. R. Abeyratne, "Blockchain ready manufacturing supply chain using distributed ledger," International Journal of Research in Engineering and Technology, vol. 5, no. 9, pp. 1-10, 2016.

[25] J. F Calzadilla, A. Villa, "Systematic Literature Review of the use of Blockchain in Supply Chain.," in 12th European Research Seminar (ERS) On Logistics and SCM, Barcelona, 2017.

[26] D. &. F. T. Bhowmik, "The multimedia blockchain: A distributed and tamper-proof media transaction framework.," in 22nd International Conference on Digital Signal Processing (DSP). , 2017.

[27] Zibin Zheng ; Shaoan Xie ; Hongning Dai ; Xiangping Chen ; Huaimin Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends.," in IEEE International Congress on Big Data (BigData Congress)., 2017.

[28] M. S. Henriques and N. K. Vernekar, "Using symmetric and asymmetric cryptography to secure communication between devices in IoT", Proc. Int. Conf. IoT Appl. (ICIOT), pp. 1-4, May 2017.

[29] T. N. a. H. Hartenstein, "Network Layer Aspects of Permissionless Blockchains," IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 838-857, 2019.

[30] C. F. Martin Harrigan, "The Unreasonable Effectiveness of Address Clustering," in IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress, 2016.

[31] Elli Androulaki, Ghassan O. Karame, Marc Roeschlin, Tobias Scherer, Srdjan Capkun, "Evaluating User Privacy in Bitcoin," in Financial Cryptography and Data Security, Japan, Springer, 2013, pp. 34-51.

[32] Malte Möser*, Kyle Soska, Ethan Heilman, Kevin Lee, Henry Heffan, Shashvat Srivastava, Kyle Hogan, Jason Hennessey, Andrew Miller, Arvind Narayanan, and Nicolas Christin, "An Empirical Analysis of Traceability in the Monero Blockchain," Proceedings on Privacy Enhancing Technologies, vol. 2018, no. 3, p. 2018, 143–163.

[33] P. Koshy, D. Koshy, P McDaniel, "An analysis of anonymity in Bitcoin using P2P network traffic," in International Conference on Financial Cryptography and Data Security., 2014.

[34] A. Biryukov, D. Khovratovich, and I. Pustogarov, "Deanonymisation of clients in Bitcoin P2P network," in ACM Conference on Computer and Communications Security, 2014.

[35] A. Biryukov, Tikhomirov, Sergei, "Deanonymization and linkability of cryptocurrency transactions based on network analysis," Proceedings of 2019 IEEE European Symposium on Security and Privacy (EuroS&P), 2019.

[36] D. Nick, Data-Driven De-Anonymization in Bitcoin, Zurich: Swiss Federal Institute of Technology, 2015.

[37] Dorit Ron, Adi Shamir, "Quantitative Analysis of the Full Bitcoin Transaction Graph," in Financial Cryptography and Data Security, SpringerLink, 2013, pp. 6-24.

[38] S. Goldfeder, H. Kalodner, D. Reisman, and A. Narayanan, "When the cookie meets the blockchain: Privacy risks of Web payments via cryptocurrencies," Proceedings of Privacy Enhancing Technoly, vol. 2018, no. 4, pp. 179-199, 2018.

[39] Mercer, N. T. Courtois and R., "Stealth address and key managementtechniques in blockchain systems," in International Conference on Information Systems Security and Privacy, Porto, 2017.

[40] Amrit Kumar, Clement Fischer, Shruti Tapole, Prateek Saxena, "A Traceability Analysis of Monero's Blockchain," in European Symposium on Research in COmputer Security, Oslo, Norway, 2017.

[41] Chaum, D.L.,, "Untraceable electronic mail, return addresses, and digital pseudonyms.," Communication ACM, vol. 24, no. 2, p. 84–90., 1981.

[42] G. Maxwell., "CoinJoin: Bitcoin Privacy for the Real World.," Bitcoin Forum , 2013. [Online]. Available: https://bitcointalk.org/index.php?topic=279249.

[43] Jan Henrik Ziegeldorf, Fred Grossmann, Martin Henze, Nicolas Inden, and Klaus Wehrle. , "Coinparty: Secure multi-party mixing of bitcoins.," in Proceedings of the 5th ACM Conference on Data and Application Security and Privacy,, 2015.

[44] Licheng Wang, Xiaoying Shen, Jing Li, Jun Shao, Yixian Yang, "Cryptographic primitives in blockchain," Journal of Network and Computer Applications, vol. 127, pp. 43-58, 2019.

[45] R. Rivest, A. Shamir, and Y. Tauman. , "How to leak a secret," Asiacrypt 2001, vol. 2248, p. 552–565, 2001.

[46] Yifan Wu. , An E-voting System based on Blockchain and Ring Signature, University of Birmingham, 2017.

[47] Liu, J.K., Wei, V.K., Wong, D.S. , "Linkable spontaneous anonymous group signature for ad hoc groups (extended abstract). In: Information Security and Privacy," in 9th Australasian Conference, ACISP , Sydney, Australia, 2004.

[48] Fujisaki, E., Suzuki, K., " Traceable ring signature. In: Public Key Cryptography," in 10th International Conference on Practice and Theory in Public-key Cryptography, Beijing, China, 2007.

[49] Back, A.,, "Bitcoins with Homomorphic Value (Validatable but Encrypted).," 2015. [Online]. Available: https:// bitsharestalk.org/index.php/topic,16797.msg214814.html#msg214814.

[50] N. van Saberhagen, "Cryptonote V 2.0.," 2013. . [Online]. Available: https://cryptonote.org/whitepaper.pdf. [Accessed 11 December 2019].

[51] Maxwell, G., "Confidential transactions.," 2017. [Online]. Available: https://people.xiph.org/greg/confidential_ values.txt.. [Accessed 11 December 2019].

[52] Danny Yang Jack Gavigan Zooko Wilcox'O'Hearn, "Survey of confidentiality and privacy preserving technologies for blockchains," [Online]. Available: https://z.cash/static/R3_Confidentiality_and_Privacy_Report.pdf.

[53] Noether, S., "Ring Signature Confidential Transactions for Monero," IACR Cryptology , p. 1098, 2015.

[54] S Goldwasser, S Micali, and C Rackoff., "The Knowledge Complexity of Interactive Proof-systems.," 1985.

[55] Miers, I., Garman, C., Green, M., Rubin, A.D., , "Zerocoin: anonymous distributed e-cash from bitcoin," in IEEE Symposium on Security and Privacy, Berkeley, CA, USA, 2013.

[56] Ben-Sasson, E.,Chiesa, A.,Garman, C.,Green, M., Miers, I.,Tromer, E., Virza,M., "Zerocash: decentralized anonymous payments from bitcoin," in IEEE Symposium on Security and Privac, Berkeley, CA, USA, 2014.

[57] George Kappos, Haaroon Yousaf, Mary Maller, Sarah MeikleJohn, "An Empirical Analysis of Anonymity in Zcash," in 27th USENIX Security Symposium., USA, 2018.

[58] Seres, István András et al. , "MixEth: efficient, trustless coin mixing service for Ethereum," IACR Cryptology ePrint Archive, 2019.

[59] G. 2. Maxwell, "Coinswap: Transaction Graph Disjoint Trustless Trading.," October 2013.

[60] Ruffing, T., Moreno-Sanchez, P., Kate, A.,, "Coinshuffle: practical decentralized coin mixing for bitcoin.," in European Symposium on Research in Computer Security, 2014.