Survey on Blockchain Privacy Challenges

Constance Hendrix and Rory Lewis

University of Colorado, Colorado Springs CO 80919, USA

Abstract. Blockchain is the underlying technology behind cryptocurrency and is migrating quickly to other industry applications. Given the rapid growth of Bitcoin and Ethereum, advances in blockchain has proliferated; however, privacy threats still linger. This persistent concern has spawned continuing analysis of existing and emerging threats and innovative pathways for solutions to blockchain in cryptocurrency and smart contracts. In addition, the critical issue of increasing the scalability of blockchain without trading decentralization and security continues to challenge researchers. Herein, we present a concise analysis of the blockchain process, detail categorical privacy vulnerabilities with notable key solutions, discuss the emergence of oracle systems, then highlight promising directions for future research.

Keywords: Blockchain · Privacy · Oracles.

1 Introduction

Blockchain is the foundational concept behind the cryptocurrency trendsetter, Bitcoin. Its debut was made by Satoshi Nakamoto, whose work built upon concepts introduced as early as the 1980s [13]. Since then, cryptocurrency options have become forefront in the debate to replace fiat currencies, blockchain technology has transitioned to other industries, and advancements have been made with security; however, problems with privacy, scalability, digital wallets, smart contract, decentralized applications (dApps), and exchange security still exist [36] [16] [28]. In industries where personal identifiable information (PII) is utilized, and the risk of identity theft is viable, it is imperative that privacy be prioritized. The threat is not limited to the blockchain itself, but off-chain storage and external systems which interface with oracles to support smart contract execution. Given this, additional vulnerabilities exist in terms of transaction privacy and reputation manipulation. In this paper, we perform a literature review which starts broad, inspecting the area of privacy, then focus promising research directions, which will inform the scope of and targets for future testing.

2 Blockchain Basics

The core of Bitcoin public blockchain is its ledger, which is openly distributed via a peer-to-peer (P2P) network. Because P2P consensus is central to the system, disrupting data integrity is difficult. A crypto transaction starts when an end

user *i)* creates a digital wallet [22], and *ii)* digitally signs a transaction to send to the *blockchain*. The blockchain is, in essence, a digital ledger that keeps track of each crypto transaction and as it is duplicated and distributed across the entire network of computer systems, some of which are operated by *miners* [38]. Miners collect the transactions, add them to a block, then determine the block's hash puzzle solution, known as a *nonce*, which is their *Proof of Work* (PoW) [8]. While the first miner publishes the block, which includes the nonce, block number, previous block hash, timestamp, and merkle root hash, other miners may follow by publishing a block containing these same transactions, potentially creating multiple forks. Herein, timeliness is critical and having a trusted clock is exceedingly important [25]. The miner who successfully attributes the most cumulative work to the blockchain wins the reward and transaction fee [35], while the forks created by other miners will be orphaned. For Bitcoin, consensus within the P2P network follows a democratic validation schema and occurs every 10 minutes, solidifying the group's resulting determination. Essentially, all nodes, including miners, have a unifying view of the blockchain and can view and verify all ongoing changes. The ledger uses hash functions to create fixed length hashes given any arbitrary length of information which provide a check of the integrity in the published block. Hash functions such as `SHA256`, are publicly known, but deriving information strictly from its hash is computationally non-trivial. Each block not only contains its own hash, it also contains the previous block's hash, hash pointer, along with the Merkle root hash. The `SHA256` hash used by Bitcoin is of the *Merkle Damgård Construction*, which pads the information, to create the root hash, divides the hash into blocks using the hash function $f : 0, 1^{|bits|}$, then compresses the data to a fixed length. In contrast, `KECCAK256`, used by Ethereum, is of the *Sponge* construction, which works via *absorb-and-squeeze-operations* on the data. Regardless of what algorithm is used, the retrieving validated information is always embedded in the block and is dependent upon *i)* the nonce discovery, *ii)* the hash of the previous block, and *iii)* the Merkle root determination. Given that a nonce is b length, each miner is mandated to check all 2^b combinations until a solution is found that allows the unlocking of the information. The inclusion of the previous block's hash provides the link in the chain, while the Merkle Tree method generates a Merkle root hash of all hashes within the block which validates the block and improves blockchain scalability [22]. The essence of a Merkle tree as it travels over four transactions within a block can be modeled as, $H_{mr} = H\left(H(H(m_1)||H(m_2))||(H(m_3)||H(m_4)\right)$, where H_{mr} is the Merkle root hash, H is the hash function, $||$ indicates concatenation, and m indicates the message contained therein [27]. As blocks are published using Bitcoin, despite the identity being somewhat protected, the details of the transaction are made available in various public blockchain implementations such as `"blockchain.com/explorer"`.

3 Privacy Vulnerabilities

The debate between privacy versus the attribution for illegal activities such as tax evasion [16] and the now extinct Silk Road [28] continues. Regardless, the need to protect personal identity and activities, which is sub-categorical to "security", remains relevant. In the US, Nolo's Plain-English Law Dictionary defines privacy as the "the right not to have one's personal matters disclosed or publicized; the right to be left alone," [17] while European Commission Regulation 2018/1725 addresses privacy through the protection of personal data. These two referenced understandings provide the scope of privacy for this paper: protection of individual's activities, correspondence, identity, and personal records. Given this scope, four areas within the field of blockchain are investigated, with the takeaways provided up front:

1. Private keys - key discovery due to poorly randomized key construction [12] or using quantum computers [23]
2. Transactions - identity discovery through network analysis of transaction patterns using network analysis and behavior-based clustering [40]
3. Personal Records - compromise of personal identifiable information [34]
4. Smart Contracts - protection of user activity due to malicious manipulation of external data sources (i.e., oracles) [28]

3.1 Private Keys

Asymmetric encryption is primarily used with cryptocurrency systems such as Bitcoin and Ethereum; it is also used with private and consortium blockchains. Zhang *et al.*, proposed an e-Health system using a public key encryption in a private blockchain [41]. Other applications relying on asymmetric encryption to ensure privacy include but are not limited to voting systems[33], crowdsourcing systems [29], and information retrieval systems [4], which is why preventing key compromise is essential. Additionally, digital wallet privacy breeches can result in lost assets [36] when the end user's private key is lost in emails and or when user names are physically lost. Three other approaches to protect these keys are to use: 1) an additional layer of security on the user's device; 2) biometrics to prevent unauthorized access to the key; and, 3) biometrics in "known cryptography algorithms" [9]. However, it is worth mentioning that biometrics is also considered identifiable data, therfore should be protected as well. In addition, [16] highlighted companies similar to Chainalysis, Elliptic, and DMG, whose business is to discover a user's identity through digital wallet addresses. Keeping wallets offline instead of online also reduces the risk of key compromise. The key algorithm used by Bitcoin and Ethereum is the 256-bit Elliptic Curve Digital Signature Algorithm (ECDSA), specifically `secp256k1` [12]. ECDSA's *Elliptic Curve* enhancement is used with more foundational algorithms to reduce computing power, which is favorable for implementation on mobile devices [39]. Although ECDSA seems to be secure, vulnerabilities do exist [3]. The Schnorr signature, which unlike ECDSA, is linear and more conducive to applications

such as *i)* the Naïve Signature Aggregation where a user's private key is part of a collective signature used ultimately to sign a transaction, and *ii)* trusted external data feeds, supplied by oracle systems later discussed. However, it is vulnerable to compromise using a rouge key attack [31]. MULTISIGNATURES, which allow a group of users to sign a single document and may have alleviated the Schnorr vulnaribility by leveraging key interaction or challenges [10], [31].

3.2 Transactions

The goal of transaction privacy is to protect privacy of transaction contents such as time, currency amount, and addresses from unauthorized entities [19]. To counter, a popular method for increasing transaction transparency is to analyze patterns within the blockchain network by using behavior-based clustering techniques, such as k-means, then providing cluster definitions and rules in order to characterize human behavior in the post-analysis. Supervised learning in conjunction with k-means has been used to more accurately define clusters [6]. Silhouette scores have also been used with clustering to determine sufficiency of each object's classification [19]. In addition, other methods used to compromise transaction security include transaction "fingerprints", pattern determination, network traffic analysis using mass data collection, and transaction propagation techniques [19] [26]. For example Biryukov *et al.*, performed a transaction propagation analysis to link transactions and a method "for linking transaction clusters to IP addresses of [initiating] nodes" [11]. Additionally, transaction propagation within a system is often defined by the software used to conduct cryptocurrency transactions and ledger updates such as advertisement-based propagation, send-headers propagation, unsolicited push propagation, relay network propagation, and push/advertisement hybrid propagation [28].

In light of these efforts, many solutions for protecting transaction privacy have been devised [23]; which includes mixing and anonymous solutions [24], [19]. *Mixing* obfuscates transactions by mixing and re-distributing, while *Anonymous* removes transaction payment origins. Other solutions comparable are also available. For example, the system *Hawk* addresses the issue by providing end users the capability to privately create and interact with smart contracts using zero knowledge proof protocols, then further protects data by storing transactional data off-chain [25].

Block synchronization may also differ between systems. In 2015, Bitcoin updated its broadcasting protocol from "trickle" to "diffusion" spreading propagation protocol which defined the delays between transaction transmission to the nodes, neglecting to significantly improve the lack of anonymity in its network, as in the case of [18]. In the event of system updates or attacks, such as smart contract hack on *Decentralized Autonomous Organization* (DAO) venture capital fund [37], ledger inconsistencies between nodes can be introduced. The *Proof of Communication* consensus-based solution, controls timing and claims to be a more secure option over PoW and *Proof of Stake* (PoS), later discussed [15].

Government entities are also targeting law breakers. For example in effort to enforce tax laws, the US's Internal Revenue Service has a history of subpoenaing

companies owning cryptocurrency system [5] and hiring others equivalent to *Chainalysis* to assist in the investigation of illegal activity. Michael Gronager, CEO of *Chainalysis*, stated *Chainalysis* "builds personas around the transaction patterns, then attributes them to entities" [14]. Although *Chainalysis* is arguably a means for for a safer tomorrow, compromise of privacy is the price.

3.3 Personal Records

Although there is no precedence for *PII* in public blockchains, private and consortium, quasi-private, blockchain implementations may include identification authentication to ensure access to information, and transaction initiation is more controlled. *PII* could also be included as transactional data, which include, but are not limited to, biometrics, medical records, or government issued identification. Compromise of identity due to authentication is covered in [34]. In addition to authentication, *PII* will most likely be included in e-Health records and systems supporting real estate, insurance, public benefit, and voting management. Although the restricted access or private and consortium blockchains are more efficient compared to the public, and minimum privilege is a traditional security strategy, there are privacy risks associated with these options [42].

3.4 Smart Contracts

The *Smart Contract*, initially proposed by Nick Szabo in 1994, was first integrated to support digital currency by Ethereum. A smart contract used on blockchain with dAPP, minus the front-end user interface, is executable code that is task dependent on external data from a trusted sources called *oracles* and other pre-defined conditions. Ethereum, being the first to use this concept on a public blockchain, relies on programming languages such as *Solidit*y or *Vyper* to create smart contracts to be executed on the *Ethereum Virtual Machine* (EVM). After the code is compiled and executed on EVM, it is broadcasted to all nodes with access where end-users execute contracts by submitting a transaction. Although Ethereum was the first to implement smart contracts, smart contracts are also used by others like *HyperLedger*, which uses a peer-designated verification system and secures chaincode in a *Docker* container. For contracts in general, privacy concerns creep into security of contract data fields designated as "private", data source authenticity, and vulnerabilities of trusted data feeds [7] [32] [19]. Public blockchains housing these contracts maintain secrecy by using cryptographic techniques, heavily relying on the robustness of the keys or rules, but may be compromised depending on transaction behaviors [7]. Using external data in contracts while maintaining privacy is arguably more challenging. Oracle schemes, such as voting-based systems such as *Chainlink*, provide this service but doesn't actually authenticate the data nor invoke Transport Layers Security (TLS) using third party verification [32]. Although the following section elaborates this service, other proposed solutions, ranging in maturity, are noteworthy: *Ziraffe*, *Enigma Secret Contracts* [43], and *Town Crier* [23].,

4 Oracle Systems Emergence

To expound, *oracles* are systems used to facilitate the use of trusted external data in contracts and are provided to many blockchains, to include Bitcoin, Ethereum, and HyperLedger. Consider ChainLink, which is comprised of a decentralized oracle system residing on Ethereum, whose functions can be divided into one of two categories: on-chain and off-chain. On-chain is the label assigned to those functions which have a direct interface with the blockchain ledger, whereas off-chain functions do not. Data Providers, such as *Binance*, *GamesScoreKeeper*, and *Amberdata*, are external to the system, but provide necessary external data needed by the system for the service to work. To explain, Fig. 1 steps through ChainLink's process, depicting the interactions within and between [1]. The on-chain *ChainLink Smart Contract* acts as the interface between the blockchain's smart contract and the off-chain ChainLink node, routing requests and data. Recently, ChainLink announced its movement from using *FluxAggregator* for aggregating data on-chain to off-chain, improving scalability of increasing data needs [21] and reducing gas costs related to publishing data on Ethereum. Coined by ChainLink, "*Off-Chain Reporting*" (OCR) reports and digitally signs observations into a single report, then sends it to the chain for a smart contract signature verification [20]. To prevent data providers from sending false data to the system, its trust model requires the provider to submit a stake in their native LINK token, Chainlink's proprietary Ethereum token. If data provided is deemed truthful, a reward is administered; if not, a penalty to the stake is applied.

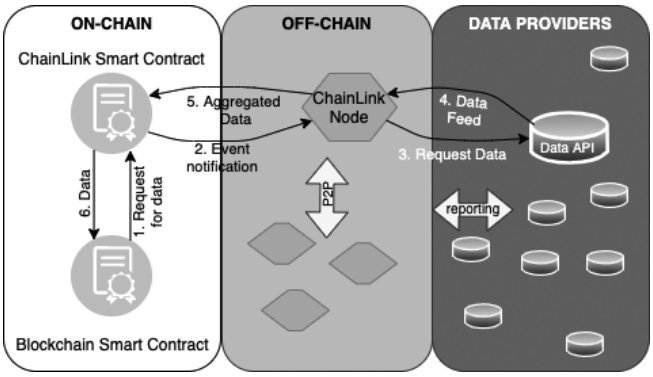

Fig. 1. ChainLink High-Level Functional Diagram is divided into three layers.

A specific oracle trust model defines its process to increase privacy protection of external data. However, the risk of corrupted external data feeds, reputation manipulation such as a Sybil attack, collusion between oracles, or identity thefts exist [2]. Al-Breiki *et al* identified a plurality of trust models, where systems are grouped and linked to its adopted trust model. These models are categorized as on-chain, off-chain, and on-/off-chain, a hybrid approach. On-going research is being conducted to improve existing methods, but challenges still exist. [32] [2].

5 Conclusion

Although cryptocurrencies are successful in contributing to value exchange, they still have to overcome the distrust in the process from those who believe it represents another tulip bubble [30]. Regardless, markets leveraging blockchain are growing and applications involving different industry sectors, to include edge computing, are taking hold. Therefore, establishing robust privacy and security techniques and practices should be a priority in future design. Recent surveys have covered a variety of topics from blockchain malicious attacks to solutions in disparate industries. However, surveys on private and consortium blockchain privacy issues, papers providing detailed comparison of oracle security in current implementations, and techniques determined to be state-of-the-art were more of a challenge to locate. In this paper, we provided an overview of blockchain, then investigated privacy vulnerabilities within four areas. In future work, we plan to focus our efforts to create privacy-centric, scalable solutions to address corrupted external data feeds and reputation manipulation attacks targeting oracle systems, while identifying current state-of-the art methods and open challenges.

References

1. Blockchain oracles for connected smart contracts | chainlink, `https://chain.link/`
2. Al-Breiki, H., et al.: Trustworthy blockchain oracles: review, comparison, and open research challenges. IEEE Access **8** (2020)
3. Aldaya, A., et al.: Port contention for fun and profit. In: 2019 IEEE Symposium on Security and Privacy. pp. 870–887 (2019)
4. Amiri, W., et al.: Privacy-preserving smart parking sys using blockchain and private information retrieval. In: 2019 International Conference on SmartNets (2019)
5. Aquillo, M.: Court grants IRS summons of Coinbase records (2018), `https://www.journalofaccountancy.com/issues/2018/mar/irs-summons-of-coinbase-records.html`
6. Aspembitova, A., et al.: Behavioral structure of users in cryptocurrency market. PLOS ONE **16** (2021)
7. Atzei, N., et al.: A survey of attacks on Ethereum smart contracts. In: Principles of Security and Trust. pp. 164–186. Springer (2017)
8. Aura, T., et al.: Dos-resistant authentication with client puzzles. In: International workshop on security protocols. pp. 170–177. Springer (2000)
9. Aydar, M., et al.: Private key encryption and recovery in blockchain. arXiv:1907.04156 [cs] (2020)
10. Bellare, M., Neven, G.: Multi-signatures in the plain public-key model and a general forking lemma. In: ACM Computer and Comm Security Proceedings (2006)
11. Biryukov, A., Tikhomirov, S.: Deanonymization and linkability of cryptocurrency transactions based on network analysis. In: 2019 IEEE European Symposium on Security and Privacy. pp. 172–184. IEEE Xplore (2019)
12. Breitner, J., Heninger, N.: Biased nonce sense: Lattice attacks against weak ECDSA signatures in cryptocurrencies. In: Financial Cryptography and Data Security. pp. 3–20. Springer (2019)
13. Buterin, V.: A next generation smart contract and decentralized application platform (2014), `https://cryptorating.eu/whitepapers/Ethereum/Ethereum_white_paper.pdf`

14. CB Insights: Chainalysis (2019), https://www.youtube.com/watch?v=yNpNz-FvSYQ
15. Chen, Y., et al.: DEPLEST: A blockchain-based privacy-preserving distributed database toward user behaviors in social networks. Information Sciences (2019)
16. Dasgupta, D., et al.: A survey of blockchain from security perspective. Journal of Banking and Financial Technology **3**, 1–17 (2019)
17. Editors, N.: (2021), https://www.law.cornell.edu/wex/right_to_privacy
18. Fanti, G., Viswanath, P.: Anonymity properties Bitcoin P2P network. arXiv (2017)
19. Feng, Q., et al.: A survey on privacy protection in blockchain system. Journal of Network and Computer Applications **126**, 45–58 (2019)
20. Foxley, W.: Off-chain reporting, https://docs.chain.link/docs/off-chain-reporting#how-does-it-work, assessed 2021-03-14
21. Foxley, W.: Chainlink promises 10x data with new off-chain reporting overhaul (2021), https://www.nasdaq.com/articles/chainlink-promises-10x-data-with-new-off-chain-reporting-overhaul-2021-02-24
22. Gupta, S.S.: Blockchain. IBM Onlone (http://www. IBM. COM) (2017)
23. Hasanova, H., et al.: A survey on blockchain cybersecurity vulnerabilities and possible countermeasures. International Journal of Network Management **29** (2019)
24. Joshi, A.P., Han, M., Wang, Y.: A survey on security and privacy issues of blockchain technology. Mathematical Foundations of Computing **1**, 121 (2018)
25. Kosba, A., et al.: Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. In: IEEE symposium on security and privacy (2016)
26. Koshy, P., et al.: An analysis of anonymity in bitcoin using P2P network traffic. In: International Conference on Financial Cryptography and Data Security (2014)
27. Krzyzanowski, P.: Week 9: Blockchains & Bitcoin (2020), https://www.cs.rutgers.edu/~pxk/419/notes/pdf/09-bitcoin-slides.pdf
28. Li, X., et al.: A survey on the security of blockchain systems. Future Generation Computer Systems **107**, 841–853 (2020)
29. Lin, C., et al.: SecBCS: a secure and privacy-preserving blockchain-based crowdsourcing system. Science China Information Sciences **63**, 1–14 (2020)
30. Liu, Y., Tsyvinski, A.: Risks and returns of cryptocurrency. Tech. rep., National Bureau of Economic Research (2018)
31. Maxwell, G., et al.: Simple Schnorr multi-signatures with applications to Bitcoin. Designs, Codes and Cryptography **87**, 2139–2164 (2019)
32. Park, J., et al.: Smart contract data feed framework for privacy-preserving oracle system on blockchain. Computers **10**, 7 (2021)
33. Pawlak, M., et al.: Voting process with blockchain technology. In: Advances in Intelligent Networking and Collaborative Systems. pp. 233–244 (2019)
34. Rana, R., et al.: An assessment of blockchain identity solutions: Minimizing risk and liability of authentication. In: 2019 IEEE WIC ACM International Conference on Web Intelligence (WI). pp. 26–33. IEEE Xplore (2019)
35. Seguias, B.: To fork or not to fork: the blockchain's propensity to converge (2018), https://delfr.com/wp-content/uploads/2018/09/Blockchain_Forks.pdf
36. Spadafora, A.: Blockchain hacks led to billions in losses last year (2021), https://www.techradar.com/news/blockchain-hacks-led-to-billions-in-losses-last-year
37. Tikhomirov, S., et al.: SmartCheck: static analysis of ethereum smart contracts. In: Proceedings of the International Workshop on Emerging Trends in Software Engineering for Blockchain. pp. 9–16. ACM (2018)
38. Wüst, K., Gervais, A.: Do you need a blockchain? In: 2018 Crypto Valley Conference on Blockchain Technology (2018)

39. Yassein, M.B., et al.: Comprehensive study of symmetric key and asymmetric key encryption algorithms. In: International Conference on Eng and Tech). pp. 1–7. IEEE Xplore (2017)
40. Yu, T., Cao, C.: Privacy protection in blockchain systems: A review. In: Data Processing Techniques and Applications for Cyber-Physical Systems. pp. 2045–2052. Springer (2020)
41. Zhang, A., Lin, X.: Towards secure and privacy-preserving data sharing in e-Health system via consortium blockchain. Journal of Medical Systems **42**, 140 (2018)
42. Zheng, Z., et al.: Blockchain challenges and opportunities: a survey. International Journal of Web and Grid Services **14**, 352–375 (2018)
43. Zyskind, G., et al.: Enigma: Decentralized computation with guaranteed privacy. arXiv (2015)