Blockchain vs GDPR in Collaborative Data Governance

Rahul Dutta[1], Arijit Das[2], Ayan Dey[1], and Sukriti Bhattacharya[3(✉)]

[1] University of Calcutta, Kolkata, India
rahul39dutta@gmail.com, adakc_rs@caluniv.ac.in
[2] Narula Institute of Technology, Kolkata, India
arijit1080@gmail.com
[3] Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg
sukriti.bhattacharya@list.lu

Abstract. Data Governance is the trending topic in today's security-privacy-concerned digital ecosystem. Blockchain technology is probably one of the most acclaimed evolutions in recent times. Blockchain technologies can be a game-changer for data governance in the areas of transparency and data provenance. As a distributed ledger technology (DLT), blockchain is being touted as a potentially transformational force in collaborative data governance. The General Data Protection Regulation (GDPR) entered into force on May 25, 2018. It is the latest in a series of European Union (EU) legislative measures designed to give EU citizens more control over their data. GDPR, which directs a centralized 'data controller' (GDPR Article 4) to manage user data, clashes with the blockchain's decentralized data storage and management process. The GDPR and the blockchain both have a common ideological ground, emphasizing the need for a change in managing personal data. While GDPR takes care of the policy side by setting up a standard, the blockchain helps enable the implementation side by providing a unique framework. In this paper, the authors analyze the clashes between the two and the potential solutions to those clashes for blockchain to comply with GDPR.

Keywords: Data governance · GDPR · Blockchain · Data protection

1 Introduction

Collaborative data governance is the future of data management practices in an organization to keep pace with the dynamic nature of today's business world. On the broad spectrum, it answers the following questions related to the data: the data owner, the data quality, how to manage the data, and the possible use cases for the data. The world became aware of blockchain technology in 2009 with the release of Bitcoin [1] by Satoshi Nakamoto, a digital currency that is independent of a central authority. Blockchain technology is versatile enough to handle any information that can be digitized beyond the cryptocurrency. Blockchain holds

Y. Luo (Ed.): CDVE 2020, LNCS 12341, pp. 81–92, 2020.
https://doi.org/10.1007/978-3-030-60816-3_10

built-in security and privacy. It offers trust, transparency, autonomy, and consensus. The collaborative data governance allots with the issues of availability, usability, security, data integrity, and analytics. Blockchain has suddenly surfaced as an elixir for these issues with a throng of facilities and benefits that enhance data governance and increase trust. Blockchain simplifies the management of trusted information over communication networks by enabling transparent interactions among different parties in a faster, safer, and more reliable way. Given that the blockchain is trustworthy, secure, and cannot be tampered with, it is seeping into corporate awareness. From what data governance needs and what blockchain provides, it's easy to see they are an excellent fit for each other.

In parallel, the General Data Protection Regulation, better known by its acronym, GDPR is a new framework for consumer data protection that came into effect in Europe from 25^{th} May 2018. The implementation of the GDPR is fundamentally linked to a company's data governance program. It significantly empowers several rights where 'data subjects'[1] can demand companies to provide them with the whereabouts of their personal data and what processing is being done on them. The regulation addresses data protection rules for personal data export outside of the EU and enforces EU data protection laws to guide foreign organizations that process personal data about residents of the EU (GDPR Chapter 5)[2].

In some critical ways, blockchain shares many goals with the GDPR. However, there exist some real conflicts between GDPR and blockchain as well. In this paper, we discuss these issues from the collaborative data governance perspective. The rest of the paper is divided into three Sections - **GDPR against the Blockchain** (Sect. 3), **GDPR for the Blockchain** (Sect. 4) and **Conclusion** (Sect. 5).

2 Background

In this Section General Data Protection Regulation[3] and blockchain [1] is explained in brief.

2.1 General Data Protection Regulation (GDPR)

Ensuring the confidentiality of personal data of data subjects and giving them the option to impose access, erasure and rectification of the data are the primary goals of GDPR. The GDPR realizes four different roles : (a) Data Subject, the provider of the data to the controller (GDPR Article 4(1)), (b) Data Controller, the legal body who determines the purposes of data and the means of their processing (GDPR Article 4(7)), (c) Data Processor, the legal body that processes the data as determined by the data controller (GDPR Article 4(8))

[1] GDPR Article 4(1), https://gdpr-info.eu/article-4.
[2] GDPR Chapter 5, https://gdpr-info.eu/chapter-5/.
[3] GDPR, www.gdpr-info.eu/.

and (d) Third Parties, who are not the above three, but have the authority of data processing under the authority of the controller (GDPR Article 4(10)).

GDPR gives "privacy by design" significant importance. It calls for inclusion of data protection from the onset of designing systems, implementing appropriate technical and infrastructure measures (GDPR Article 25)[4]. Failing to adhere to the regulation will result in a fine of upto 20 million Euros or 4% (or 2%) of global turnover, whichever is higher[5].

2.2 Blockchain

A blockchain is a network that uses distributed ledger technology (DLT) where the data stored is decentralized, transparent and immutable. Data is stored in a list of blocks where each block, added in an append-only manner contains a hash of previous block, a time stamp, and a set of data as shown in Fig. 1.

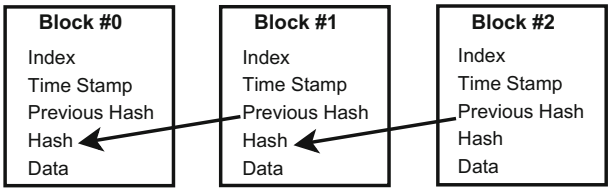

Fig. 1. Blocks in a blockchain

The key properties of blockchain in terms of the three types of blockchain are shown in Table 1.

Table 1. Types of blockchain and their properties

Properties	Public permissionless	Public permissioned	Private Permissioned
Data control	Decentralized	Decentralized	Decentralized
Network	Highly Decentralized	Semi-centralized	Highly centralized
Privacy	None	Low	High
Border	Cross-bordered	Cross-bordered	Bordered
Data immutability	Yes	Yes	Yes
Data persistency	Yes	Yes	Yes
Anonymity	High	Low	None

[4] GDPR Article 25, www.gdpr-info.eu/art-25-gdpr/.
[5] GDPR Fines/Penalties, https://gdpr-info.eu/issues/fines-penalties/.

3 GDPR Against Blockchain

In this section, we investigate the significant conflicts in the GDPR-blockchain relationship and the legal challenges for blockchain to comply with the regulation. The conflicts and challenges are discussed by answering the following five questions.

Q3.1: How does GDPR's right to be forgotten (Article 17) and right to rectification (Article 16) clashes with blockchain's immutable ledger?

□ GDPR right to be forgotten, also called right to erasure directs that the data subjects (personal data subjects) shall have the "right to obtain from the controller the erasure of personal data concerning him or her without undue delay" as stated in GDPR Article 17 (1). The right is not absolute and applies only in certain circumstances such as if the personal data is no longer in use "for the purposes for which they were collected" (GDPR Article 17(1(a))) or the data subject withdraws consent to processing of the data (GDPR Article 6 and Article 9)[6] [7] and their is "no other legal ground for the processing" (GDPR Article 17(1(b))) or in circumstances where their is 'unlawful' processing of the personal data (GDPR Article 17(1(d))). If the controller has made the personal data public, "reasonable measures, including technical measures" must be taken taking into account available technology and the cost of implementing such measures (GDPR Article 17(2)). However, blockchain is fundamentally immutable. A typical transaction in a blockchain, consists of the sender, recipient and amount of asset transfered which are personal data and are stored permanently on the ledger. It is designed to retain its information indefinitely as a mechanism to eliminate any fraud and assures ownership of data in the ledger. Moreover, GDPR addresses data adjustability in right to rectification by mandating that the data subject can exercise their "right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her" (GDPR Article 16). Any modification of the data on the blockchain breaks the links between the blocks and therefore, the chain is broken and the entire blockchain is rendered useless. GDPR's demand for data minimization is also another blow to the blockchain's immutable nature. According to the regulation, the personal data must be "adequate, relevant and limited" to protect privacy so that only absolutely necessary facts about the data is shared (GDPR Article 5(1(c)))[8].

[6] GDPR Article 6, https://gdpr-info.eu/art-6-gdpr/.
[7] GDPR Article 9, https://gdpr-info.eu/art-9-gdpr/.
[8] GDPR Article 5, https://gdpr-info.eu/art-5-gdpr/.

Q3.2: How GDPR's right to restriction of processing (Article 18) clashes with blockchain's distributed ledger?

□ GDPR defines data processing as "any operation or set of operations which is performed on personal data or on a set of personal data, whether or not by automated means" (GDPR Article 4(1)). GDPR incurs greater data control power to data subjects by giving them the right to restriction of processing, that is, "The data subject shall have the right to obtain from the controller restriction of processing" (GDPR Article 18) if the data subject feels the data may be inaccurate or there is "unlawful" processing of the personal data. However, the data as a part of the ledger in a blockchain network is distributed amongst the nodes participating in the network. Besides the data controller, every participant in the network can process the data which averts data privacy desired by the regulation. In simple terms, every participant maintains a local copy of the ledger. Anyone can store and process the data, because of this local copy held in his or her computer, while maintaining their anonymity. This directly contradicts GDPR's privacy policies where a centralized data controller and processor has the power to store, collect and process the data. (GDPR Article 24)[9].

Q3.3: Can encrypted personal data be stored on a blockchain? Are encrypted personal data on a blockchain anonymous enough to fall outside the scope of GDPR?

□ GDPR does not apply to anonymous data as stated in GDPR Recital 26[10]. The threshold for anonymization under the regulation is high and only results "from processing personal data in order to prevent identification" irreversibly [2]. A typical blockchain network aims to achieve anonymity using cryptography, precisely, public-key cryptography and hashing. However, the identity and data in a blockchain is in fact, 'pseudonymous'. According to Andreas M. Antonopoulos[11], Bitcoin is mistakenly characterized as 'anonymous' currency [3]. Identity is tied to a pseudonym, that is, the public key. Pseudonymous data, however, falls under the scope of the law. Data is pseudonymous if it can be tied to other available information to identify the data subject, else the data is considered anonymous (GDPR recital 26).

Q3.4: In a blockchain who is identified as data controller and data processor?

□ According to GDPR Article 13, data subjects have the right to know "the identity and the contact details of the controller"[12]. GDPR Article 4 (7) and Article 4 (8) define that in order to process data, data controllers are required

[9] GDPR Article 24, https://gdpr-info.eu/art-24-gdpr/.
[10] GDPR Recital 26, https://gdpr-info.eu/recitals/no-26/.
[11] https://antonopoulos.com/.
[12] GDPR Article 13, https://gdpr-info.eu/art-13-gdpr/.

to determine "the purposes and means of the processing of personal data" and data processors are required to process "personal data on behalf of the controller". GDPR's accountability directs a controller to be responsible for legal and lawful processing of personal data (GDPR Article 5). A *public permissionless* blockchain is perfectly decentralized as there is no inherent power inequality between the nodes. Thus, *permissionless* blockchain networks require no obvious identifiable data controller to direct and co-ordinate the working of the blockchain network and it falls directly under the scope of GDPR. In a *permissioned* blockchain network, identity of the data controller is less of an issue. GDPR Article 26(1)[13] also defines joint controllers "where two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers". In a permissioned blockchain, a plurality of controllers can govern the network with pre-defined rules of consensus, which is exactly how a *federated* blockchain network works.

Q3.5: How does GDPR's territorial scope clashes with blockchain?

□ According to GDPR Article 3 territorial scope, the regulation applies to personal data processing of the citizens who are a part of EU. In GDPR Chap. 5, "Transfer of personal data to third countries or international organization", the law defines that personal data transfer to countries and controllers outside EU are allowed only if the controller conforms to GDPR or provides an identical level of data protection (GDPR Article 45(1))[14] and where "enforceable data subject rights and effective legal remedies for data subjects are available". The regulation applies to controllers who are not established in the EU but process personal data of the citizens in EU. *Public* blockchains do not provide adherence to these principles and data movement in them are cross-bordered. In a *public* blockchain, anyone can join the network as a node and create transactions. *Public* blockchain provides no geographical border to its network, which comprises a number of uncontrolled nodes, free to trade data and thus the allows cross-border data processing. Compliance with GDPR's territorial scope is more feasible in a *private permissioned* blockchain.

4 Towards GDPR Compliant Blockchain

This Section states the potential solutions to the conflicts addressed in Sect. 3 between GDRP and blockchain in data driven cooperation.

Q4.1: How to enable GDPR's right to erasure and rectification in a blockchain?

□ We look at the potential solutions to enable data modification in blockchain.

[13] GDPR Article 26, https://gdpr-info.eu/art-26-gdpr/.
[14] GDPR Article 45, https://gdpr-info.eu/art-45-gdpr/.

Storing Data Off-Chain. A potential solution to allow data adjustability in blockchain is storing the personal data off-chain with restricted access control. The reference to this data is kept along with its hash and other metadata including the claims and permissions regarding this data on the blockchain [4]. The blockchain is a medium of enabling trust between parties and delivering proof for transactions. The proof will be done off-chain through conventional methods where the data controller decides if the party requesting the data has authorized access to it [5]. After GDPR came into play, off-chain data storage in the blockchain is now seen as a possible workaround.

A number of off-chain data storage solutions for blockchain have emerged recently. My Health My Data (MHMD), a *private permissioned* blockchain[15], is a project funded by the EU, stores personal data off-chain. Other such blockchain networks include the Bitcoin Lightning Network [6] that uses off-chain exchange of bitcoins between nodes through micropayment channels[16]. Æternity blockchain [7] uses state channel technology[17] for transactions and smart contracts to be executed off-chain. Liquidity exchange built on top of Liquidity network [8] allows off-chain exchanges, which although being a decentralized network is scalable to a centralized exchange.

State Tree Pruning and Smart Contract Self-destruct in Ethereum. State tree pruning[18], similar to automatic memory management in volatile resources, is a method of removing data in the Ethereum blockchain. However, the drawback is that it is intended for minimizing states in the block by removing unused records and removal does not depend on participant's demand. The only way to remove code from the blockchain is when a contract at that address performs the "selfdestruct" operation[19]. The storage and code is removed from the state. Even if a contract is removed by "selfdestruct", it is still part of the history of the blockchain and probably retained by most Ethereum nodes. So using "selfdestruct" is not the same as deleting data from a hard disk.

Chameleon Hashes. Using "chameleon" hash functions [9] to frame an editable blockchain (or redactable blockchain) has been explored [10]. The hash of the data stored in the blockchain retains the integrity of the data. Any changes to the data changes the hash of the data in the block which implies data tempering. This changes the block header hash and eventually breaks the chain (link) between the subsequent blocks. Also referred to as trapdoor hash functions, "chameleon" hash function has an additional secured private key, known as trapdoor key using which the original data can be updated and written back

[15] www.myhealthmydata.eu/.
[16] https://medium.com/@super3/introduction-to-micropayment-channels-5beb3bb224c1.
[17] https://blog.stephantual.com/what-are-state-channels-32a81f7accab.
[18] https://blog.ethereum.org/2015/06/26/state-tree-pruning/.
[19] https://solidity.readthedocs.io/en/v0.5.0/introduction-to-smart-contracts.html?highlight=self%20destruct.

into the blockchain. However, the updated data has the same hash value as the original data. This enables imposing GDPR's right to erasure and right to rectification by allowing users to rewrite or delete past blocks of information without breaking the blockchain.

μchain. Recently, authors have introduced a new mutable blockchain, called μchain [11]. The main features of μchain include maintaining alternative versions of data records, using consensus to approve a valid history and its inherent capability to conceal alternative versions of history. In a given set of transactions, only a single transaction is marked as "active", while all the remaining ones represent alternative inactive transactions. A set of transactions can be extended to add new versions of transactions. Thus, transactions can be updated by the sender if it requires any rectification which is the "active" transaction. Decryption keys are available only for the "active" transactions, and the inactive transactions are all kept hidden by encryption. This allows data subjects with the option to exercise their right to rectification.

Q4.2: How to identify data controller and data processor in a blockchain?

☐ GDPR is established on the notion of an identifiable data controller to determine "the purposes and means of the processing" (GDPR Recital 79)[20]. Here, we address how can such a controller and processor be established in the decentralized world of blockchain.

Network Administrator and Miners in `Permissioned` Blockchain. In a *permissioned* blockchain, the entity responsible for handling of personal data can be the data controller and data processor depending on the implementation specifics. The entity who determines the participation rights (authentication) and the access rights (authorization) of every participants in the blockchain network can be considered as the controller of the network. This entity can be the governing body, that is, "the natural or legal person" which "determines the purposes and means of the processing of personal data" (GDPR Article 4(7)). In a *federated* blockchain, a plurality of controllers can govern the network with predefined and transparent rules of consensus, that is, in case of joint controllers. The miners make a paramount contribution to the operation of the blockchain as they bear the responsibility of confirming and adding transactions into the blockchain. The transactions initiated by the participants in a blockchain are stored in a temporary 'pool of unconfirmed transactions' which is a collection of transactions waiting for their confirmation [3], that is, waiting to be processed by the miners. Therefore, we can consider miners as the data processors. Hence, in accordance with GDPR data controller definition (GDPR Article 4(7)), they cannot be reckoned as data controllers.

[20] GDPR Recital 79, https://gdpr-info.eu/recitals/no-79/.

Smart Contract. A blockchain network employing smart contracts protocol can operate it as a data processor. CNIL[21] explained in their report[22] that smart contract developers could be considered either data controllers or data processors accordingly. For a specific purpose and means for processing of personal data for that purpose identified by a participant or set of participants in the blockchain, a developer designs a smart contract solution for that purpose.

Here, the smart contract developer is processing the personal data on behalf of that participant through that smart contract. Here, that participant can be considered as the data controller (or joint controllers in case of plurality of controllers) whereas the smart contract developer as well as the smart contract itself can be determined as the data processor.

Q4.3: How to make personal data anonymous in a blockchain?

□ Recalling GDPR Recital 26, anonymous data are not in the scope of the regulation. If the data on the blockchain is anonymous, it falls directly outside the scope of GDPR. In this Section, potential solutions to achieve anonymity are discussed.

Zero-Knowledge Proof. A solution to achieve anonymity in blockchain is Zero-knowledge proof [12]. It allows a verifier to validate the truth of a statement of a party (prover) without having any knowledge about the statement except that the statement is true. Zero-knowledge protocols supplements privacy in transactions in a blockchain where the transaction is validated without revealing the internal details of the transactions, that is, the sender, recipient and the content of the transaction (personal data). The entire blockchain network can agree on the validity of a transaction without learning about the content of the transaction.

Zero-knowledge protocols in blockchain use zk-snarks (Zero knowledge Succinct Non-Interactive Arguments of Kn-owledge) [13] to overcome multiple challenge interactions between verifier and prover to validate a transaction. Zk-snarks only requires a common string of characters known by the prover and verifier to validate a statement. Using the digital signature of the prover in blockchain as the common string, a transaction can be proved easily with low computational effort. The personal data in the transactions are decentralized and no personal data is stored on the blockchain, Due to the absence of actual data in the blockchain, participants cannot process the data which in turn, provides data privacy and restricts any unintended and "unlawful" processing of the data as desired in GDPR's right to restriction of processing as explained in Sect. 3.

[21] https://www.cnil.fr/en/home.

[22] CNIL report,https://www.cnil.fr/fr/Blockchain-et-rgpd-quelles-solutions-pour-un-usage-responsable-en-presence-de-donnees-personnelles.

Zcash[23] is a high profile cryptocurrency using zk-snarks. Ethereum's metropolis[24] protocol upgrade allowed users to use zk-snarks in their smart contracts. PIVX is another cryptocurrency using zk-snarks[25].

Ring Signatures and Stealth Addresses. Another solution to anonymization in blockchain is the use of ring signatures [14] and stealth addresses. Ring signatures hides the identity of the creator/sender of a transaction. A ring signature is a type of digital signature in which a group of possible signers are fused together to obscure the actual signature of the sender of the transaction. All the signers are equal and valid and hence, to a third party all the signers seem to be the sender but is unable to determine the identity of the actual sender of the transaction. Stealth addresses, enhances privacy by shielding the identity of the recipient in a transaction. A transaction in blockchain is essentially transformation of old outputs belonging to one wallet to new outputs in another wallet. Stealth address is recorded in every transaction to indicate the recipient of asset involved. However, any other party cannot determine the identity of the recipient by looking at the stealth address because a stealth address is not associated with the recipient's public address. Stealth addresses and ring signatures are used in Bytecoin[26] and Monero[27]. Monero also uses the technique of ring confidential signatures [15], which is a combination of ring signatures, to hide transaction amounts.

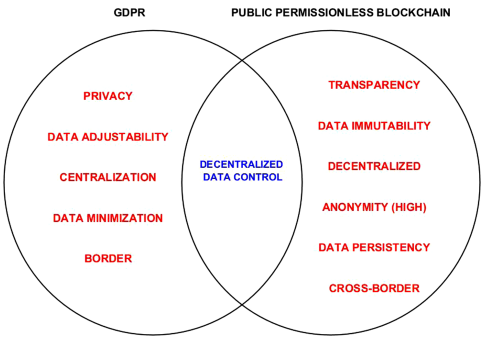

Fig. 2. GDPR vs *Public permissionless* blockchain

[23] Zcash, 2018, www.z.cash.
[24] https://blockgeeks.com/guides/ethereum-metropolis/.
[25] https://pivx.org/wp-content/uploads/2018/10/PIVX-White.pdf.
[26] Bytecoin, https://bytecoin.org/.
[27] Monero, https://ww.getmonero.org/.

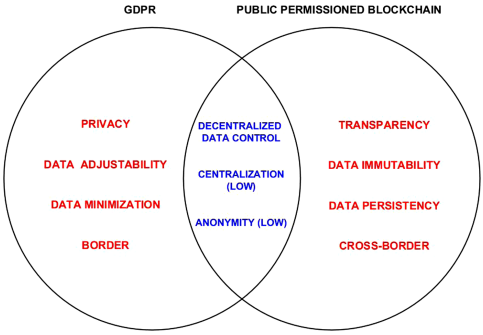

Fig. 3. GDPR vs *Public permissioned* blockchain

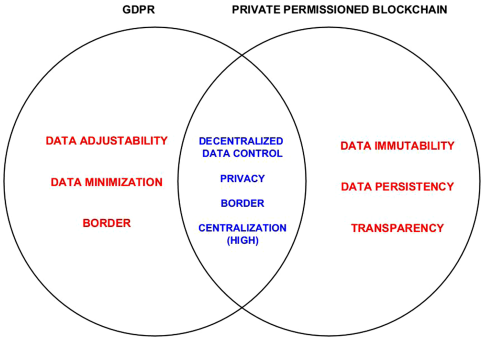

Fig. 4. GDPR vs *Private permissioned* blockchain

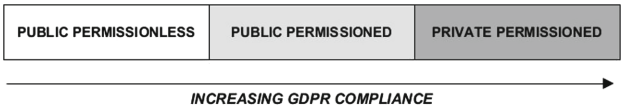

Fig. 5. GDPR compliance

5 Conclusion

This paper analyzed and discussed the significant privacy concerns and challenges to blockchain from the lens of collaborative data governance; precisely the rights of data subject, data controller and processor, pseudonymous personal data and the territorial scope of GDPR. The paper also highlights opportunities for the various types of blockchain to comply with GDPR or at least achieve partial compliance which itself is a step forward. As a result, we get a clear vision of the relationship between GDPR and *public* and *private* blockchain networks respectively. We illustrate their relationship through venn diagrams. Fig. 2 shows the relationship between GDPR and *public permissionless* blockchain. Fig. 3 and Fig. 4 illustrate the relationship between GDPR and *public permissioned*

blockchain and *private permissioned* blockchain respectively. From the above comparison, we get a clear understanding of the potentials of GDPR compliance of the three different types of blockchain networks as shown in Fig. 5. We can observe from Fig. 5 that the increase in GDPR compliance is when the decentralization nature of the blockchain decreases. To demonstrate the compatibility of blockchain and GDPR, the possible solution discussed in this paper should be leveraged to the most significant extent possible in blockchain solution architectures.

References

1. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system. Technical report (2008)
2. Finck, M.: Blockchains and data protection in the European union. Technical report 18-01 (2018)
3. Antonopoulos, A.M.: Mastering Bitcoin: Unlocking Digital Crypto-Currencies, 1st edn. O'Reilly Media Inc, Newton (2014)
4. Humbeeck, A.V.: The blockchain-GDPR paradox, Medium (2017)
5. Xu, X., et al.: A taxonomy of blockchain-based systems for architecture design, April 2017
6. Poon, J., Dryja, T.: The bitcoin lightning network: scalable off-chain instant payments (2016). https://lightning.network/lightning-network-paper.pdf
7. Hess, Z., Malahov, Y., Pettersson, J.: Æternity blockchain (2017)
8. Stiller, B., Bocek, T., Hecht, F., Machado, G., Racz, P., Waldburger, M.: Mobile Systems IV, University of Zurich, Department of Informatics, Technical report, January 2010
9. Krawczyk, H.M., Rabin, T.D.: Chameleon hashing and signatures, US Patent 6,108,783, August 22 2000
10. Ateniese, G., Magri, B., Venturi, D., Andrade, E.: Redactable blockchain-or-rewriting history in bitcoin and friends. In: 2017 IEEE European Symposium on Security and Privacy (EuroS&P), pp. 111–126. IEEE (2017)
11. Puddu, I., Dmitrienko, A., Capkun, S.: μchain: how to forget without hard forks (2017)
12. Goldwasser, S., Micali, S., Rackoff, C.: The knowledge complexity of interactive proof systems. SIAM J. Comput. **18**(1), 186–208 (1989)
13. Bitansky, N., Canetti, R., Chiesa, A., Tromer, E.: From extractable collision resistance to succinct non-interactive arguments of knowledge, and back again. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, ser. ITCS 2012. New York, NY, USA, pp. 326–349. ACM (2012). http://doi.acm.org/10.1145/2090236.2090263
14. Rivest, R.L., Shamir, A., Tauman, Y.: How to Leak a Secret. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 552–565. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45682-1_32
15. Noether, S., Mackenzie, A., et al.: Ring confidential transactions. Ledger **1**, 1–18 (2016)