Implementing Blockchain Technology to Comply with GDPR Legislation, a Simple Solution

Andy Ho, An Nguyen and Jodi Pafford

***Abstract*— Blockchain is a highly attractive new technology, due to its security features, in a time where concern for consumer data protection is on the rise. The immutability of blockchain through its decentralized framework is an appealing option for conforming to regulations protecting consumer data privacy such as the European Union's (EU) General Data Protection Regulation (GDPR) and similar legislation proposed in the United States (US), however companies needing to conform to regulations will likely need to keep costs as low as possible and solutions simple. Although many other proposals have introduced the idea of a blockchain-database hybrid or a mutable blockchain, this paper seeks to implement a third, more simple solution.**

I. Introduction

Blockchain technology has attracted interests from a wide span of industries; mainly due to its ability to operate in a decentralized fashion [1]. A blockchain utilizes a digital ledger of transactions where all participants edit in a secure way and is shared over a distributed network of computers [2]. The blockchain has an append-only structure, which helps it protect old data against modification or deletion [3]. In order to make changes, all the nodes present in the network must evaluate, verify, and match the transaction information; if the majority of the nodes agree a new block is added to the chain [2]. Considering several recent highly publicized data breaches raising public concern, and in the face of looming legislative changes in the United States [4, 5], after the implementation of the General Data Protection Regulation (GDPR) [6], this extra-secure framework is an attractive one.

Some possible challenges for companies trying to conform to GDPR-like legislation include the "Right to be forgotten" where citizens are given strict control over their personal data [7]. Conforming to a GDPR-like regulation and securing the data are two dynamically opposed paradigms that must be reconciled with each other.

One solution that has been presented includes implementing a blockchain-traditional database hybrid where user data is stored on a traditional database and modifications are recorded on the blockchain [8, 7]. Alternatively, another solution is a blockchain with the ability to forget has been proposed. This proposal uses the pruning features of traditional blockchains like Bitcoin or newer smart-contracts like Ethereum to remove blocks in a traceable way [3].

In this paper, we will give a brief overview of the blockchain and its background. We will give a condense summary of the European Union's GDPR. Following this, we will discuss the different methods proposed and their weaknesses. The remainder of this paper will contain our proposal, its implementation and drawback.

II. Blockchain

Blockchain technology has the potential to redesign how computational resources interact in an automated and decentralized society. This technology was invented by a person (or group of persons) named Satoshi Nakamoto in 2008 to be used as a public transaction ledger for a cryptocurrency called Bitcoin [9]. On a high-level, the ledger is a self-governing list of records called blocks [9] which are linked cryptographically using a hash algorithm. Each block is then connected to the previous block containing a timestamp and the transaction data. Here's a high-level diagram to illustrate the transaction flow:

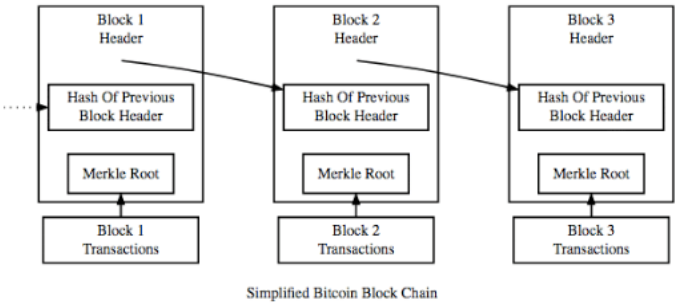

Figure 1. Blocks linked to one another

How does Blockchain Works?

The following will use Bitcoin, a cryptocurrency, to explain how blockchain technology works – it is important to note that this framework can be used in non-finance related applications as well. Owner A wants to send his Bitcoin to Owner B – which in reality assigns Owner B's identification to that specific transacted Bitcoin. For this to take place, the

transaction is inserted as a 'block' which is then broadcasted to the peer-to-peer network for verification. If successful, the transaction will be recorded in a public ledger. These 'blocks' are all linked to one another (hence, Blockchain) in a linear and chronological sequence with every block containing the hash of the previous block [10] – shown in Figure 1. Refer the following diagram so visualize the flow of how a transaction is processed for Bitcoin:

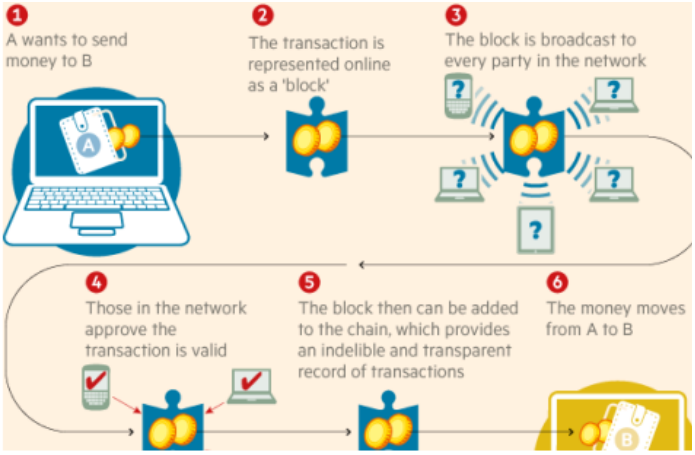

Figure 2. Flow chart of Bitcoin Transaction

What is a Cryptographic Hash Function?

A cryptographic hash function protects sensitive information, either at rest or in transit. In the case of Blockchain, it prevents the double-spend problem (definition: the act of using the same coin more than once) from occurring with the use of public-key cryptography [11]. A transaction is first initiated by future owner of the cryptocurrency by sending his public key to the current owner. The cryptocurrency is then transferred by the digital signature of a hash – the public keys (i.e. the assigned address of the cryptocurrency) are stored in the blockchain. In the case of Bitcoin, it utilizes a SHA-256 hash function which take an input of a random size and produces an output of a fixed size – pre-image resistant. What makes SHA-256 powerful is that it is nearly computationally infeasible to reconstruct a given input from the output value.

Immutability of Blockchain

Immutability is one of the most important (and defining) features of blockchain. As explored in the section above, blockchain is a one-way hash function making the records irreversible without community consensus. This eliminates reconciliations and establishes trust in the system [12]. Only the owner of the record who has the proper credentials could make changes to the records.

Industry Applications of Blockchain

Since 2008, blockchain technology has been considered for many use cases outside of cryptocurrency (i.e. finance related). The following are industries where blockchain technology could be applied:

1. Insurance: With the use of a blockchain application called 'Smart-Contracts (reference), insurance claims could be processed without the aid of an adjuster or physical inspection. Users would provide the information and the smart-contract would determine if it satisfies the criterions prior to distributing funds to the insured.
2. Internet of Things (IoT): There are new security vulnerabilities since these devices are sending and receiving data. Blockchain would add an additional layer of security insuring only the owner is receiving access and information of his IoT devices.
3. Healthcare: The encryption blockchain provides would be of importance to medical records, prescriptions, and supply management – offering extreme privacy.
4. Voting: Voter fraud could be prevented since a vote recorded in the blockchain would be immutable and would provide an audit trail if there is evidence of tampering. Since each voter would have his own lock and key, authentication of user would virtually be impossible to fake.

Although created for the financial world, the implications of Blockchain technology can affect applications in wide range of areas outside of finances. One of the main purposes of blockchain is to eliminate the need for an intermediary to verify and process a transaction – making it decentralized and distributed. The main objective is that the blockchain establishes a new standard and model by creating a distributed consensus in a digital world [10]. In order to accomplish this, the designers made blockchain immutable and/or resistant to data modification – any change would require all subsequent blocks to be altered. Since these public ledgers are managed by peer-to-peer network [10], mass collaboration governs whether a transaction can be changed. This headwind hits directly at one of the central tenants of GDPR [13].

As one would expect, the supporters of blockchain believe the advantages of it outweigh the regulatory issues – the inverse applies for those who hold data privacy in higher regards. Before moving on to evaluating possible solutions that can satisfy both worlds, lets dive deep into GDPR.

III. General Data Protection Regulation

What is GDPR?

"The improvement in substance is that there's far more transparency under the new rules, which means that you will have more detailed information policies about what your data are processed for, which purposes if they are given to others, and there will be also in general more possibilities to get a view of which data are there about you. And you have new rights like data portability and the right to be forgotten. So, it will

be far easier for consumers to control their personal data."

- Jan Philipp Albrecht, member of the European Parliament and 'father' of the GDPR [13]

GDPR was created from the European Commission to reform data protection across the European Union in order to make Europe 'fit for the digital age' [14]. The origins of what is now known as the GDPR began in 2012. All organizations in the member-states across Europe, including those who have dealings with businesses in Europe must adhere to the GDPR EU framework. The GDPR was approved and adopted in April 2016 but was not enforced until May 25, 2018 [15].

The key changes in the reform include:

· *A **single set of rules** on data protection, valid across the EU.*
*Unnecessary **administrative requirements,** such as notification requirements for companies, will be removed. This will save businesses around €2.3 billion a year.*

· *Instead of the current obligation of all companies to notify all data protection activities to data protection supervisors – a requirement that has led to unnecessary paperwork and costs businesses €130 million per year, the Regulation provides for increased **responsibility and accountability** for those processing personal data.*

· *For example, companies and organisations must notify the national supervisory authority of serious **data breaches** as soon as possible (if feasible within 24 hours).*

· *Organisations will only have to deal with a **single national data protection authority** in the EU country where they have their main establishment. Likewise, people can refer to the **data protection authority** in their country, even when their data is processed by a company based outside the EU. Wherever **consent** is required for data to be processed, it is clarified that it has to be given explicitly, rather than assumed.*

· *People will have easier **access to their own data** and be able to **transfer personal data** from one service provider to another more easily (right to data portability). This will improve competition among services.*

· *A **'right to be forgotten'** will help people better manage data protection risks online: people will be able to delete their data if there are no legitimate grounds for retaining it.*

· *EU rules must apply if personal data is **handled abroad** by companies that are active in the EU market and offer their services to EU citizens.*

· ***Independent national data protection authorities** will be strengthened so they can better enforce the EU rules at home. They will be empowered to fine companies that violate EU data protection rules. This can lead to penalties of up to €1 million or up to 2% of the global annual turnover of a company.*

· *A new **Directive** will apply general data protection principles and rules for **police and judicial cooperation** in criminal matters. The rules will apply to both domestic and cross-border transfers of data* [14].

The GDPR at its core is powerful and needed, however, the application of it interferes with the premise behind Blockchain Technology. In 2012 when the European Commission first introduced the GDPR, blockchain was not a known word and the GDPR idea was initially focused on cloud services and social networks.

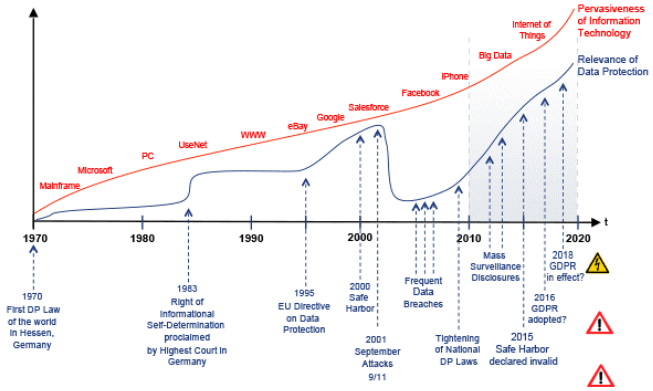

Figure 3. A brief history of the General Data Protection Regulation by Wilhelm (2016) [16]

Implications of GDPR for Blockchain

The table below, summarized from The Journal of The British Blockchain Association, summarizes the implications related to blockchain and GDPR. [13]

GDPR Article/Recital	Implications	Topic
Art. 4(1), 6(4), 32/Rec. 26	Can PD be stored on a blockchain or must be off-chain? The connection between pseudonymized and anonymized data and the data subject.	**Personal data on the blockchain**
Art. 6	Six reasons can be used to comply with lawful processing, and a data sharing agreement can be recorded on a BC	**Lawful Processing in the EU/Consent**
Art. 17, 17(1), (a,b), 6(1)(b,f)/Rec. 69	Can data on a blockchain be deleted in accordance to the RTBF and what would happen if not – could the functioning principle take over that allows for specific interpretations of the GDPR, as BC is at its core designed not to be compliant to the RTBF.	**Right to be forgotten (RTBF) and functioning principle**
Art. 25/Rec. 78	BC runs counter to data minimization, storage limitations and a clearly determined data controller, raising the question whether it is in line with 'Privacy by Design' (PbD). Privacy risks of entire IT-architecture, including BC. Solutions could be Enigma or differential privacy or future more secure BCs.	**Privacy by Design versus blockchain core features**
Art. 26(1)/Rec. 79	Private versus public BC and the accountability of a (joint) data controller.	**Accountability of data controller**

Table 1. Implications related to Blockchain and GDPR.

Article 4
Article 4 of the GDPR defines personal data. The definition is very broad which complicates it's interpretation with the use of blockchain [17].

Article 6
Article 6 discusses the consent that must be given. Such consent must have already undergone a thorough academic and practical discourse [18]. This means that consent must be "freely given, specific, informed and unambiguous".

Article 17
Article 17 of the GDPR grants EU citizens the 'right to be forgotten and to data erasure' at any time upon request. Due to the immutable nature of blockchain, this presents a challenge. This is probably the largest challenge of GDPR and Blockchain.

Article 25
Article 25 of the GDPR discusses handling personal data by the concept of Privacy by Design. Privacy by design is privacy such that it "should be promoted as a default setting of every new IT system and should be built into systems from the design stage" [19]. The blockchain implication is that the data must not be stored in plaintext. GDPR does not provide many details to this and has left it up to some interpretation [18].

Article 26
Article 26 discusses the description of who is responsible. This must be completed in a transparent manner in order to be in compliance, which can be a challenge when there are joint data controllers.

IV. Related works

With the growing interest in blockchains as a information storage system and the growing concern for user data privacy research into designing a blockchain based system that are compliant with privacy laws such as the GDPR have begun. Here we identify two strategies proposed by researchers to answer the "right to be forgotten" clause of the GDPR. One proposal by Farshid et al. is to use a pruning algorithm on a smart contract based blockchain, such as Ethereum, and will be describe in the following section [3].

Smart contract based blockchains have the additional attribute of being able to execute code. All the nodes and blocks together makeup one instance of a virtual machine. This machine can store all account balances and active codes. Once deployed smart contracts only has write-access and cannot be updated or changed. Smart contract based blockchain is attractive because the virtual machine does not require a transaction history to operate but only the current state of the machine. The researchers made use of the pruning algorithm in two Ethereum implementations to delete as much state data as possible without breaking the functionality of the blockchain. Furthermore, the researchers deleted all historical blocks and logs leaving only the current state active. With these changes it was shown that a five host machines were able to form a network and perform basic transactions. They were able to show that an account can be changed and then the history of the exchange deleted. Contracts were able to be created and

then the creation transaction be deleted. Figure 1 is a before and after screenshot of an account balance and creation of a smart contract with their transaction history deleted.

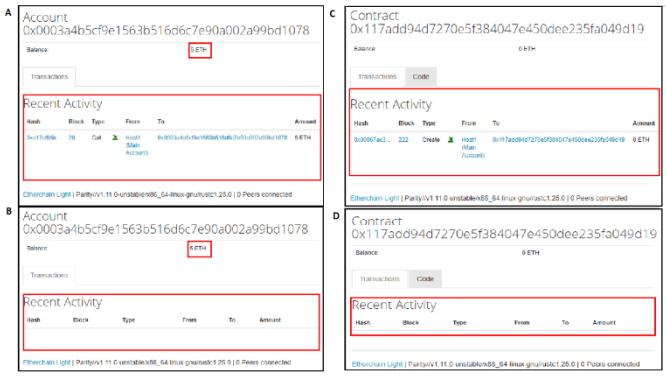

Figure 4. Demonstration of pruning algorithm

There are several limitations to this approach pointed out by the authors. One, no new nodes can be added to the network. This is because the information needed to derive the current state of the virtual machine no longer exist. Two, there is no way to prevent individuals from creating backups of old data before it gets deleted.

A different approach is described by Coelho et al [7]. The authors proposed a hybrid system where all "meaningful data" is stored off the blockchain and on a third-party database system. A digest of each instance of data and all transactions performed on the data is stored on the blockchain ledger. This way data can be deleted when requested while at the same time trust in the integrity of the data is proven with the immutability of the blockchain. Figure 5 is a diagram of such a design.

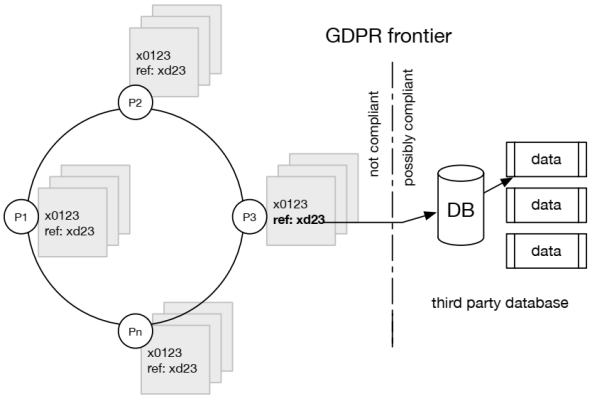

Figure 5. On-ledger/off-ledger hybrid proposed by Coelho et al.

The hybrid system proposed can be integrated into a much larger architecture that uses smart contracts to request permission from the data owners [20]. Figure 6 is a diagram of the ecosystem proposed by Faber et al. using the on-ledger/off-ledger hybrid data storage system. Briefly, the ecosystem utilizes smart contracts to store conditions for data exchanges between users and service providers as well as users and data purchaser. The required permissions allow each member of the ecosystem to interact with the blockchain, contains pointers to the actual data. A second blockchain is used to store hases of data allowing for the data purchasers to verify the integrity of the data they are accessing. Finally, the data itself is stored off-chain on third party databases.

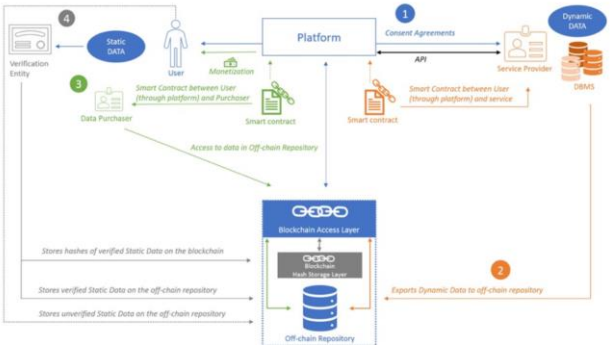

Figure 6. Personal data ecosystem proposed by Faber et al.

Using blockchains to store data is attractive in that a blockchain is immutable and that data is guaranteed to be "un-hackable." When data storage is moved off the chain it becomes GDPR compliant in that data curators can delete or modify it in accordance with the wishes of the data owner. However, this also opens up avenues for attackers to modify this data.

V. Bibliography

[1] K. Christidis and M. Devetsikiotis, "Blockchains and Smart Contracts for the Internet of Things," *IEEE Access*, vol. 4, pp. 2292 - 2303, 2016.

[2] S. Singh and N. Singh, "Blockchain: Future of Financial and Cyber Security," in *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, Noida, India, 2017.

[3] S. Farshid, A. Reitz and P. Roßbach, "Design of a Forgetting Blockchain: A Possible Way to Accomplish GDPR Compatibility," *Hawaii International Conference on System Sciences* |, pp. 7087-7095, 2019.

[4] National Conference of State Legislatures, "Data Disposal Laws," National Conference of State Legislatures, 04 January 2019. [Online]. Available: http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx. [Accessed 1 January 2019].

[5] National Conference of State Legislatures, "Security Breach Notification Laws," National Conference of State Legislatures, 29 September 2018. [Online]. Available: http://www.ncsl.org/research/telecommunications-and-information-technology/data-disposal-laws.aspx. [Accessed 31 January 2019].

[6] THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION, *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR)*, L 119 ed., 2016.

[7] F. Coelho and G. Younes, "The GDPR-Blockchain Paradox: A Work Around," in *Workshop on GDPR Compliant Systems*, Rennes, France, 2018.

[8] C. Molina-Jiménez, I. Sfyrakis, E. Solaiman, I. C. L. Ng, W. Meng, Wong, A. Chun and J. Crowcroft, "Implementation of Smart Contracts Using Hybrid Architectures with On-and Off-Blockchain Components," *ResearchGate*, pp. 1-12, 2018.

[9] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," *www.cryptovest.co.uk*, 2008.

[10] M. Crosby, Nachiappan, P. Pattanayak, S. Verma and V. Kalyanaraman, "BlockChain Technology: Beyond Bitconin," *Applied Innovation Review*, no. 2, pp. 6-19, 2016.

[11] T. Aste, P. Tasca and T. D. Matteo, "Blockchain Technologies: foreseeable impact on industry and society," *IEEE Computer*, vol. 50, no. 9, pp. 18-28, 2017.

[12] M. Pilkington, "Blockchain technology: principles and applications," in *Research Handbook on Digital Transformations*, Northampton, MA, Edward Elgar Publishing, 2016, pp. 225-253.

[13] S. Schwerin, "Blockchain and Privacy Protection in the Case of the European General Data Protection Regulation (GDPR):A Delphi Study," *The Journal of The British Blockchain Association*, vol. 1, no. 1, pp. 1-75, 2018.

[14] European Commission - Press release, *Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses*, Brussels: European Commission, 2012.

[15] Trunomi, *EU GDPR COMPLIANCE WITH TRUNOMI – ARTICLE SUMMARIES & SOLUTIONS*, GDPR, Whitepapaer, 2017.

[16] E.-O. Wilhelm, "A brief history of the General Data Protection Regulation," International Association of Privacy Professionals, 2019. [Online]. Available: https://iapp.org/resources/article/a-brief-history-of-the-general-data-protection-regulation/. [Accessed 13 3 2019].

[17] N. Kramer, "Blockchain, Personal Data and the GDPR Right to be Forgotten," 17 April 2018. [Online]. Available: https://www.blockchainandthelaw.com/2018/04/blockchain-personal-data-and-the-gdpr-right-to-be-forgotten/. [Accessed 13 March 2018].

[18] C. Wirth and M. Kolain, "Privacy by BlockChain Design: A Blockchain-enabledGDPR-compliant Approach for Handling Personal Data," *roceedings of the 1st ERCIMBlockchain Workshop 2018, Reports of the European Society for SociallyEmbedded Technologies*, pp. 2510-2591, 2018.

[19] B.-J. Koops and R. Leenes, "Privacy regulation cannot behardcoded. A critical comment on the‘privacy by design’ provision in data-protection law," *International Review of Law, Computers & Technology* , pp. 37-41, 2013.

[20] B. Faber, G. Michelet, N. Weidmann, R. R. Mukkamala and R. Vatrapu, "BPDIMS: A Blockchain-based Personal Data and Identity Management System," *Proceedings of the 52nd Hawaii Interantional Conference on System Sciences*, pp. 6855-6864, 2019.