II. The Midas touch of Blockchain: Leveraging it for Data Protection

A. Easing into the Blockchain enigma

Before we jump in to the rabbit hole that is the relationship between blockchain and the GDPR, a brief explanation of the technology is imperative. Although hailed as the disruptive technology of this century, Marco Iansiti and Karim Lakhani refer to blockchain technology as a 'foundational' model.[12] They explain that blockchain does not offer a truly 'disruptive' model in the sense that it is not capable of undercutting an existing model with a low-cost solution; rather it resonates better as a 'foundational' model by creating new foundations for social and economic purposes.[13] Drawing parallels with the adoption of TCP/IP - the distributed computer networking technology that established the foundation for the Internet - Iansiti and Lakhani highlight that it took more than 30 years to put the transformative potential of TCP/IP to use.[14] However, studies like the annual Gartner Hype Cycle (which ascertains the promise of emerging technologies) not only includes but showcases blockchain amongst the technologies capable of delivering a high degree of competitive advantage in the coming five to ten years.[15]

The broad range of applications that blockchain is presently being put to is testament to this optimistic projection. From its first application as the underlying technology for Bitcoin, blockchain has stepped out of the shadow of virtual currency and its impact now traverses beyond financial

[12] Marco Iansiti and Karim Lakhani, 'The Truth About Blockchain' (Harvard Business Review, January-February 2017) <https://hbr.org/2017/01/the-truth-about-blockchain> accessed 27 August 2017.

[13] ibid.

[14] ibid.

[15] Gartner Press Release, 'Gartner's 2016 Hype Cycle for Emerging Technologies Identifies Three Key Trends That Organizations Must Track to Gain Competitive Advantage' (August 2016) <www.gartner.com/newsroom/id/3412017> accessed 27 August 2017.

services.[16] The new vistas being explored for application of blockchain technology include, amongst others, corporate governance, democratic participation, social institutions and identity management.

The basic principles underlying blockchain technology are its structure as a distributed database, its focus on peer-to-peer transmission for communication, its potential to offer transparency through pseudonymity and irreversibility of records, and last but not the least, computational logic. At the risk of over-simplification, blockchain can be understood as a chronological database of transactions recorded by a network of computers.[17] These computers are called "nodes". When encrypted and smaller datasets known as "blocks" are organized into a linear sequence, they result in a blockchain.[18] Wright and Di Filippi explain that these blocks contain information about 'a certain number of transactions, a reference to the preceding block in a blockchain, as well as an answer to a complex mathematical puzzle, which is used to validate the data associated with that block'.[19]

Validation on a blockchain takes place by way of a digital fingerprint created through a particular hash function. A hash function is a mathematical algorithm that takes an input and transforms it to an output.[20] Therefore, a hash is a result of cryptographically transformed original information. A hash function is critical to the blockchain technology because it is extremely difficult to recreate the input data from its hash value alone.[21] Moreover, a hash function is used to map all transactions in a block,

[16] Don Tapscott and Alex Tapscott, 'The Impact of the Blockchain Goes Beyond Financial Services' (10 May 2016) <https://hbr.org/2016/05/the-impact-of-the-blockchain-goes-beyond-financial-services?referral=03759&cm_vc=rr_item_page.bottom> accessed 30 August 2017.

[17] Aaron Wright and Primavera Di Filippi, 'Decentralized Blockchain Technology and the Rise of Lex Cryptographia' (10 March 2015) <www.intgovforum.org/cms/wks2015/uploads/proposal_background_paper/SSRN-id2580664.pdf> accessed 30 August 2017.

[18] Wikipedia, 'Blocks' <https://en.bitcoin.it/wiki/Blocks> accessed 30 August 2017.

[19] Wright and Di Filippi (n 17) 7.

[20] Marc Pilkington, 'Blockchain Technology: Principles and Applications' (September 18, 2015) in F. Xavier Olleros and Majlinda Zhegu. Edward Elgar (ed.), *Research Handbook on Digital Transformations* (2016) <https://ssrn.com/abstract=2662660> accessed 30 August 2017.

[21] ibid.

whereby any differences in input data will produce different output data.[22] Every node connected to the blockchain network is able to submit and receive transactions. Furthermore, each node participating in the network has its own copy of the entire blockchain and is periodically synchronized with other nodes to ensure that nodes have the same shared database.[23] This is crucial as it provides for an exceptional degree of resilience on account of distributed storage by multiple computers (nodes) on the network.[24] Since the shared database can be recreated in its entirety, it makes the failure of a few computers on the network irrelevant.

Another key feature of blockchain technology, also described as a kind of distributed ledger technology, is consensus. In a publicly distributed ledger anyone can create a block, however what is required is a unique chain of blocks and a way to decide which blocks can be trusted. This means that in order to ascertain the legitimacy of transactions recorded into a blockchain, the network has to confirm the validity of new transactions. Therefore, a new block of data has to be added to the end of an existing blockchain only after the nodes on the network arrive at a consensus regarding the validity of the new transaction. This consensus is achieved through different voting mechanisms within a network.[25] The most common voting mechanism, also used for Bitcoin blockchains, is the Proof of Work consensus protocol, which depends on the amount of processing power donated to the network. This protocol, also known as mining, involves participating users working to solve difficult mathematical problems and publishing the solutions. Proof of Work consensus protocol uses tangible resources like computers and electricity, making it difficult for participating users/miners to pretend that they have higher mining power on the network than they actually do. The miners are rewarded with digital tokens - for example, in the case of Bitcoin blockchains they are rewarded with Bitcoins. The Proof of Work algorithms use the number and difficulty level of the solutions being found to measure how much of the network

[22] Joseph Bonneau et al., ‘Research Perspectives and Challenges for Bitcoin and Cryptocurrencies’ IEEE Security and Privacy <www.jbonneau.com/doc/BMCNKF15-IEEESP-bitcoin.pdf.> accessed 31 August 2017.

[23] Satoshi Nakomoto, ‘Bitcoin: A Peer-to-Peer Electronic Cash System’, BITCOIN.ORG 3 (2009) <https://bitcoin.org/bitcoin.pdf> accessed 31 August 2017.

[24] Wright and Di Filippi (n 17) 7.

[25] Wright and Di Filippi (n 17) 7.

agrees on the current state of the blockchain.[26] However, this implies that a Proof of Work consensus protocol demands a lot of energy and time for running these computations, making the efficiency of the protocol questionable. Once a block is added to a public blockchain upon achieving the consensus, this block can no longer be altered and the transactions it contains can be accessed and verified by every node on the network.[27] Consequently, this permanent record can be utilized to coordinate an action or verify an event with close to unimpeachable reliability, without having to trust a centralized authority's attestation to the veracity of a transaction. It appears that the confluence of individual and systemic incentives amounts to a pioneering scheme "for eliciting effort and the contribution of resources from people to conduct various record-keeping and verification activities for the public ledger".[28]

Finally, a brief explanation of the security-enhancing feature of blockchain, i.e., the encryption protocol it follows. Blockchain uses a two-step authentication process using public-key encryption. Every participant is issued a public key, which is an algorithmically generated string of numbers/letters representing the participant. This public key can be shared to enable interaction with others. The participants are also issued one/multiple private keys, each of which is also an algorithmically generated string of numbers/letters. However, it is incumbent upon the participant to keep this private key secure. A given pair of public and private keys has a mathematical relationship allowing the private key to decrypt the information encrypted using the public key. It is important to bear in mind that although participants on the network would know the public keys of other participants, the real identity of a participant can still be protected and remains unknown.[29] This ability to remain pseudo-anonymous is the high-

[26] Ethereum Stack Exchange, 'What's the difference between proof of stake and proof of work?' <https://ethereum.stackexchange.com/questions/118/whats-the-difference-between-proof-of-stake-and-proof-of-work> accessed 31 August 2017.

[27] Wright and Di Filippi (n 17) 8.

[28] David S. Evans, 'Economic Aspects of Bitcoin and Other Decentralized Public-Ledger Currency Platforms', Coase-Sandor Institute for Law & Economics, Research Paper No. 685 3 (15 April 2014) <http://dx.doi.org/10.2139/ssrn.2424516> accessed 31 August 2017.

[29] Ashurst, 'Blockchain 101: An Introductory Guide to Blockchain', Digital Economy, 20 March 2017 <www.ashurst.com/en/news-and-insights/insights/blockchain-101/> accessed 1 September 2017.

light when we view transactions on a blockchain from a data protection perspective.

It is pertinent to bear in mind that blockchain technology has been around for almost a decade and is not a static phenomenon. Introduced as the underlying technology for the virtual currency Bitcoin, its key feature was a 'public' distributed ledger as explained in the preceding part. However, in order to keep up with the vast spectrum of blockchain technology's potential applications, another variation known as private or permissioned blockchains has emerged. This development comes in light of the fact that anyone can interact with public ledgers by reading from /writing to them, however permissioned or private blockchains are suitable for applications where transaction details are sought to be kept private and not made visible to the general network and the public.[30] This variation comes with the possibility of being able to determine who can participate in the network. The mechanism for inviting new participants to the network may vary from unanimous agreement, core group acceptance, single user invitation to a more general satisfaction of pre-determined requirements.[31]

Vitalik Buterin, from the Ethereum team, writes about two possible variations of permissioned blockchains – consortium blockchains and fully private blockchains.[32] He defines a consortium blockchain as one where the 'consensus process is controlled by a pre-selected set of nodes'. For a block to be validated in a consortium blockchain, for example, a consortium consisting of 15 institutions, each of which operates a node and of these 10 must sign every block.[33] On the other hand, a 'fully private' blockchain reintroduces the very problem sought to be resolved by blockchains –centralized control by one organization. Therefore, during the course of this thesis when private or permissioned blockchains are mentioned, it refers to a hybrid between permissioned and permissionless

[30] Hossein Kakavand, Nicolette Kost de Sevres and Bart Chilton, 'The Blockchain Revolution: An Analysis of Regulation and Technology Related to Distributed Ledger Technologies' < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2849251> accessed 1 September 2017.

[31] Goldman Sachs Global Investment Research, 'Blockchain: Putting Theory into Practice' (2016) < https://www.scribd.com/doc/313839001/Profiles-in-Innovation-May-24-2016-1> accessed 1 September 2017.

[32] Vitalik Buterin, 'On Public and Private Blockchains' (7 August 2015) <https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/> accessed 1 September 2017.

[33] ibid.

blockchain –a model that continues to evolve. Further, a permissioned blockchain is preferable given its better speed and lesser requirement of computation power, making it cheaper and faster than the public blockchain alternative. Moreover, as would be clarified in the next section, when read permissions are restricted a permissioned blockchain can provide a greater level of privacy.[34]

Another variation in the technology comes by way of a shift from the Proof of Work consensus protocol to the Proof of Stake. The difference between the two lies in the fact that Proof of Stake is not about mining, rather it is about validating.[35] The participating user who seeks to validate a block must lock up some digital currency in order to be allowed to process a transaction. In this protocol, the owner of the pledged digital currency holds a financial stake in the success of the blockchain it tracks. Therefore, in Proof of Stake consensus protocol you trust the chain with the highest collateral, and the participating users have a financial stake in the correctness and validity of the blockchain at hand. Proof of Stake algorithm decides who gets to validate the block on the basis of the financial stakes involved, and the selection process also involves some randomness to avoid the risk of reverting to a centralized system.

B. *Leveraging Blockchain Technology for Personal Data Protection*

Keeping in mind the basic concepts about the working of blockchain, we can proceed to the application of blockchain technology for the purpose of protecting personal data. The proposal of using blockchain to protect personal data was made in a pioneering paper written on the topic of decentralizing privacy.[36] It questions the current models where third parties collect and control massive amounts of personal data. Finding issue with centralized organizations amassing significantly large quantities of personal and sensitive information without adequate measures to protect the said data, a proposal for decentralizing privacy is made. In light of falling trust

[34] ibid.

[35] Ethereum Stack Exchange (n 26).

[36] Guy Ziskind, Oz Nathan and Alex Sandy Pentland, 'Decentralizing Privacy: Using Blockchain to Protect Personal Data', 2015 IEEE Computer Society - IEEE CS Security and Privacy Workshops. <www.computer.org/csdl/proceedings/spw/2015/9933/00/9933a180.pdf > accessed 1 September 2017.

levels amongst data subjects, it explores the potential of blockchains to serve functions requiring trusted computing and auditability.[37] Considering that blockchain technology is structured around a network, which is evolutionary in essence, it suggests future improvements to the technology itself and a personal data management platform based on a combination of blockchain and off-blockchain storage. The approach to such a platform is rooted in privacy considerations.

The platform comprises of three entities viz., **users**, interested in using various applications offered online; **services**, providers of these applications who require processing of personal data; and **nodes**, being the entities responsible for maintaining the blockchain. The proposal relies on two assumptions viz., blockchain being tamper-proof and that the user manages her keys in a secure manner. The first assumption calls for a sufficiently large network of nodes making the consensus protocol more reliable, while the latter requires sensitivity on the part of the user to manage her keys. The protection of personal data is sought to be achieved by setting a sort of clearing-house mechanism. By way of illustration, the blockchain accepts two kinds of transactions - one used for access control management and the other for data storage and retrieval. Once the user installs an application using this proposed platform, a shared identity between the user and the service is generated along with the associated permissions and sent to the blockchain as an access control management transaction. The data collected (which could, for example, be sensor data such as location) on the device (i.e., phone or computer) operating the application is encrypted using a shared encryption key and sent to the blockchain in a storage and retrieval transaction. This transaction is further routed to an off-blockchain key-value store, which has an interface with the blockchain, retaining only a pointer (hash of the data) to the data on the public ledger. Once this is done, the service and the user can query the data using a retrieval transaction with the pointer associated to it. The blockchain kicks in to verify if the digital signature (private key) belongs either the user or the service. An additional layer of scrutiny applies for services, whereby their permissions to access the data are checked as well. The user friendly nature of the platform is buttressed by the ease with

[37] ibid.

which the user can change the permissions granted to the service including revoking access to previously stored data.[38]

A close perusal of the model articulated by Zyskin, Nathan and Pentland shows that only the user has control over her data. The public nature of the blockchain is overcome by storing only hashed pointers in it. The decentralized nature of the blockchain, along with the digitally signed transactions, ensures that an adversary cannot pose as a user.[39] Further, even if the adversary has control over one or more nodes, it can learn nothing about the raw data because it is encrypted with keys that none of the nodes possess.[40] This model leverages the distributed network feature of blockchain against the possibility of a node tampering with its local copy of data. Risk minimization is proportional to distribution and replication of data across nodes.

Finally, this paper is far-sighted in as much as it recognizes that the model in its present form only caters to storage and retrieval queries making it inefficient for processing data. Moreover, there is always the possibility of a service querying for raw data only to save it for future processing. Therefore, this thesis finds favour with an approach where a service is never allowed to observe the raw data. The technical solution mentioned by Zyskin, Nathan and Pentland, would allow a service to run computations directly on the network and obtain results.[41] It is this variation of their model that fits in with the proposal for a digital identity management platform put forth in the next chapter.

[38] ibid 2.

[39] ibid 3.

[40] ibid.

[41] ibid.