Blockchain-GDPR Privacy by Design

How Decentralized Blockchain Internet will Comply with GDPR Data Privacy

Claudio Lima, Ph.D.
Blockchain Engineering Council, BEC Co-Founder
Vice Chair IEEE Blockchain Standards

The General Data Protection Regulation (GDPR) that was recently approved for implementation in the EEUU by May 25th, 2018 [1] is already creating some controversies, when confronted with emerging Blockchain technologies, regarding what they have most in common: data privacy and protection. These are two are essential areas where Blockchain shines.

Blockchain is defined as the new Internet layer of value, adding the trinity of T's [2] (trustability, transparency and traceability) to any asset class transaction (information/data and physical goods) in the Internet that can be authenticated, validated, traced and registered in a distributed peer-to-peer (P2P) digital ledger system. These unique characteristics open up new possibilities for new services and applications, boosting today's Internet capability. Blockchain is also part of a broader scope of Distributed Ledger Technologies (DLT) and considered to be the driving force behind all latest technology advances related to cryptocurrency [3], smart contract and the Initial Coin Offering (ICO) frenzy.

However, Blockchain is more than a simple financial platform that enables Bitcoin and cryptocurrency transactions; Blockchain is becoming the underlying layer of the future of the Internet, that creates a new wave of Decentralized Applications, called DApps, that will replace most of today's centralized cloud Internet application. With Blockchain, businesses will experience a complete transformation of their current models by removing intermediaries, reducing costs and improving the trustability of the Internet, and therefore enabling a new wave of decentralized services.

At the time, the GDPR was first proposed by the European Union in 2012 [4], the Internet was only focused around the web-centric centralized cloud-based Internet. All GDPR data collection and processing for companies and individuals were considered based solely on this centralized cloud service model world. With the full introduction of P2P decentralized Internet technologies concepts around July 2017, which is the foundation of DLT/Blockchain, the GDPR model as it has been conceived is already outdated.

The first consequence of this assumption is that the enforcement of GDPR becomes a very challenging and controversial subject. From the lawmakers and regulators perspective, it requires adjustments to the GDPR framework to adapt to the new Blockchain-driven decentralized Internet technology model. In some cases, this could even threaten the broad adoption of decentralized Blockchain models if this issue is not well understood and clarified at the beginning since Blockchain and GDPR are recent trends and the market still thinks that Blockchain is only Bitcoin and cryptocurrency and not the enabler of the next generation Internet of value layer.

These GDPR new rules grant more rights to the consumer data-owner. Fortune's Global 500 companies will spend $7.8 billion to ensure they are compliant with GDPR [5] and meet this new consumer data ownership rules. The EU GDPR establishes directives on how companies can handle Personal Identification Information (PII), which can be considered as any data that can be used to identify a specific individual such as mailing and email address, phone number, social security number, driver license and so one. It can also consider user's computer IP address, login IDs, digital images, geolocation, social media posts, digital images and behavioral data as well. GDPR non-compliance can result in fines of up to 4 percent of a company's yearly revenue or 24 million dollars (whatever amount is more significant).

Particularly, Blockchain personal ID management and processes, such as KYC - Know Your Customers and others that store and process PII are critical to the design of GDPR compliant Blockchain solutions. The main challenges are related to public permissionless Blockchain technologies where, due how the blocks and Blockchain transactions are built, all information and records that enter the distributed ledgers, which are the main components of Blockchain, are publically visible, tamper-proof and immutable, which means that the data added to the public permissionless Blockchain is there forever as they are copied to any single distributed Blockchain P2P nodes, usually called "miner", working as a large distributed database.

However, the immutability of data transactions that are imprinted in the fabric of these distributed ledgers implies that one of the key principles of GDPR, Art. 17 Right to erasure ("Right to be Forgotten"), is not met by Blockchain. This principle applies when a consumer requests the data provider, called the “controller”, to erase their personal information and due to this immutability characteristics, the data cannot be removed.

However, to understand the mechanism how Blockchain works and how the data is stored in the Blockchain layer, it is necessary to introduce another two new Blockchain concepts which are cryptography and hashing. In Blockchain world all information is encrypted and hashed using specific algorithms and functions and then stored in the Blockchain distributed ledgers. Hashing is a one-way transformation of any input data to an unreadable 64 characters (SHA-256 hash algorithm) long sequence of fixed length, called hashed data [6]. Therefore, technically speaking, the consumer personal data and information, or other metadata, when stored in the Blockchain itself cannot be modified due the principle on how Blockchain works, as explained before.

However, there are alternative solutions for this challenge, which is the adoption of off-chain data storage architectures where all GDPR sensitive information and data are stored off-chain in distributed or cloud-based servers and the hashes, which is a specific encryption of this data (the reference or linkage to this data), are stored in the Blockchain layer, which serves as control pointers to these data stored off-chain. These control pointers are not the real data themselves but a pseudonymization of the original data that is stored elsewhere in another database which is not subject to the issues regarding record immutability that Blockchain provides. Analyzing the particular case of GDPR Art. 17 “Right to be Forgotten”; when the consumer requests, the service provider can then erase the “linkability” of the Blockchain hash pointer to the data located in distributed off-chain servers and this solution should work for the purpose.

In particular, GDPR Art.25 “Data Protection by Design and By Default" is the most interesting and maybe the most controversial article related to Blockchain, since anonymization techniques are also addressed by the GDPR. In Blockchain, the pseudoanonymization technique considered is hashing, as discussed above. However, there two interpretations for the pseudoanonym linkage using Blockchain - where the user’s data creates the hash of this data. The first assumption is that when this linkage is established, it is no longer considered personal since data pseudonymisation is accomplished, but not anonymization. On the other hand, there may be still need some proof or mathematical validation, based on GDPR Art.25, that off-chain data linkage using hashing might have some small possibility of being compromised by brute force attack. Besides, there are other hashing and consensus algorithm techniques that are not even considered here, like the ones that allow the owner of smart contract transaction to validate it without revealing their personal data. The conclusion that all this points out is that this issue is a moving target as the introduction of Blockchain innovation is speeding up, and a legal-technical battle lies ahead

Another challenge not yet completely addressed for the GDPR-Blockchain compliance is that due to the decentralized ownership model of the Blockchain technology, the data processors (data controller) cannot be held legally accountable as they are not clearly defined or specified since all ownership is decentralized.

A further improvement of this off-chain Blockchain-GDPR compliance architecture is to combine public Blockchain with trusted computing enclaves to enhance privacy and security levels of the Blockchain network. These enclaves are new hardware-coded trusted nodes that enable an extra layer of security and efficiency for off-chain transactions.

Personal information and identification (PII) are just one subset of what can be stored in the Blockchain. Other data types such as the public encrypted key and other types of hashed and encrypted data that points to some particular and specialized set of data, such as P2P network and nodes machine state ID and performance, consumer indexes and so one, can also be considered. Figure 1 shows this GDPR-Blockchain compliance architecture described above.

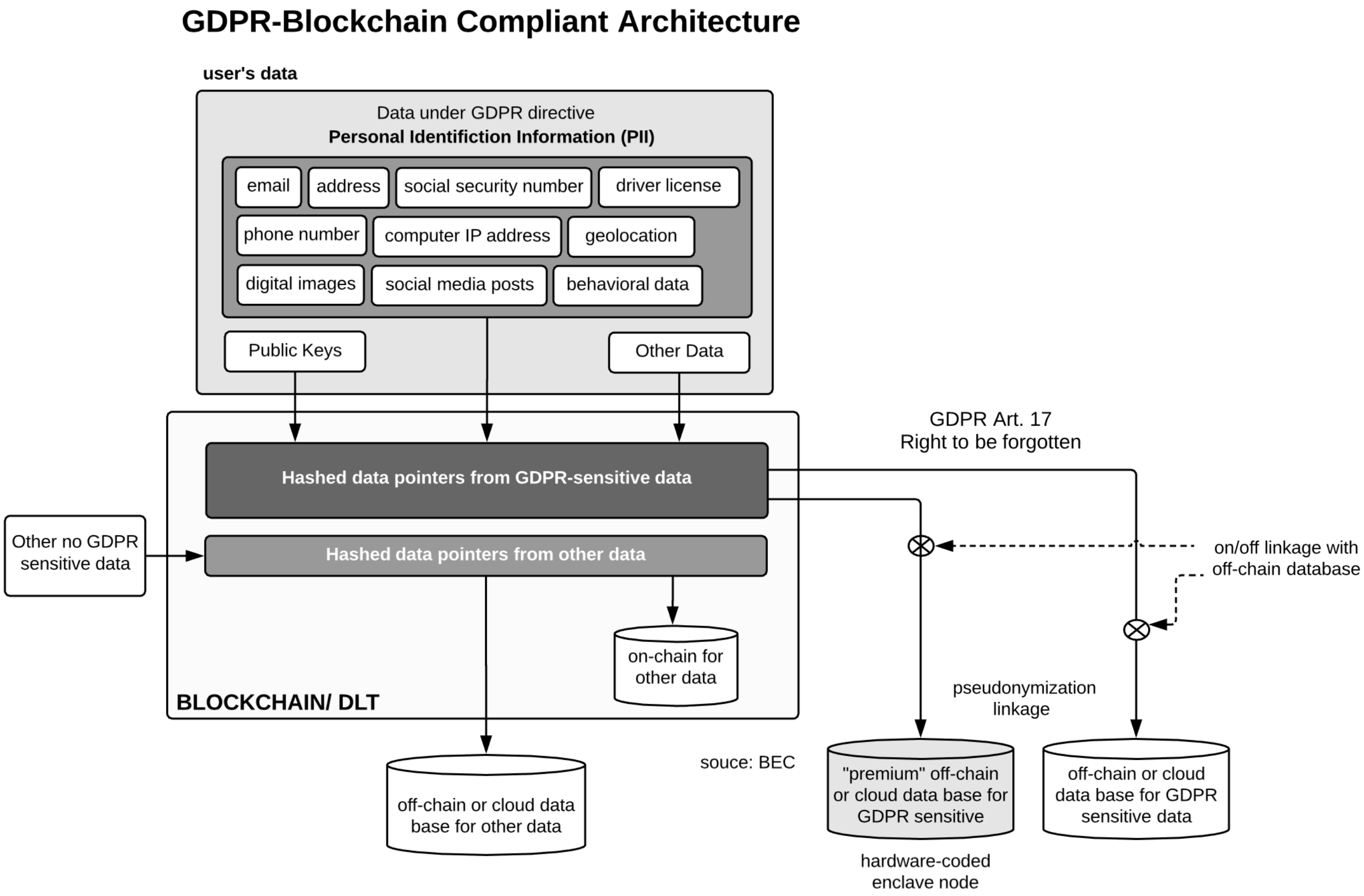
GDPR-Blockchain Compliant Architecture
user's data
Data under GDPR directive
Personal Identifiction Information (PII)
email
address
social security number
driver license
phone number
computer IP address
geolocation
digital images
social media posts
behavioral data
Public Keys
Other Data
GDPR Art. 17
Right to be forgotten
Hashed data pointers from GDPR-sensitive data
Other no GDPR sensitive data
Hashed data pointers from other data
on/off linkage with off-chain database
on-chain for other data
BLOCKCHAIN/ DLT
pseudonymization linkage
souce: BEC
off-chain or cloud data base for other data
"premium" off-chain or cloud data base for GDPR sensitive
off-chain or cloud data base for GDPR sensitive data
hardware-coded enclave node

Fig.1 - GDPR-Blockchain Compliant Architecture

There are other variants of Blockchain, called Enterprise Blockchain, that is not affected as much as the public Blockchain types are. This type of private Blockchain can easily comply with

GDPR directives since the transactions of the digital records of the stored information can be modified and erased by the private entities or authorities who can own and control this platform, using a particular class of consensus algorithm, that can handle this data accessibility, since this is a permissioned Blockchain solution.

In summary, most of the initial reaction regarding Blockchain is that this new technology is not compatible at all with the new GDPR directives. However, as it was explained here, this is a superficial assumption and conclusion to make without exploring other possibilities and clarifying how Blockchain in fact works and understanding its key underlying concepts and technologies. Indeed, Blockchain can be considered a technology that can not only improve the fundamental aspect of data privacy and security, as specified in GDPR, compared to the traditional centralized Internet approach, but can also be carefully studied, architected and implemented with a GDPR-compliance intent for data privacy, using some unique techniques. These alternatives are not simple to implement, and they require a deep understanding on how Blockchain/DLT works and how the technology ecosystem is interrelated, to create GDPR-Blockchain compatible architectures.

All the points addressed here are topics currently being debated in the industry as Blockchain-GDPR is still in the early stages of paving the way for a complete GDPR-Blockchain compliance. Even if there are technology alternatives to create GDPR compliant Blockchain systems, the debate and controversy are still alive, and the main conclusion points out for the revision of this recently introduced GDPR framework to consider the new decentralized Internet models and hash-based Blockchain technologies.

[1] General Data Protection Regulation (GDPR, https://gdpr-info.eu, May 2016.
[2] C. Lima, "Blockchain Is the New Internet - The Trinity of Ts", https://goo.gl/hq7ACy, Jan 2018.
[3] Satoshi Nakamoto. "Bitcoin: A Peer-to-Peer Electronic Cash System", bitcoin.org, 2009.
[4] Survey Fortune 500 Companies, https://goo.gl/Hqw94H
[5] The History of General Data Protection Regulation, https://goo.gl/fCWw5u
[6] Secure Hash Algorithm (SHA-256), https://en.wikipedia.org/wiki/SHA-2

Author: Claudio Lima, Ph.D.
Blockchain Engineering Council, BEC, Co-Founder
IEEE Blockchain Standards, Co-Chair
clima@blockchain-eng.org