The GDPR–Blockchain paradox: a work around

Conference Paper · December 2018

CITATIONS
0

READS
135

2 authors, including:

Fábio André Coelho
Institute for Systems and Computer Engineering of Porto (INESC Porto)
10 PUBLICATIONS **9** CITATIONS

SEE PROFILE

The GDPR-Blockchain paradox: a work around

Fábio Coelho
INESC TEC and University of Minho
Porto, Portugal
fabio.a.coelho@inesctec.pt

George Younes
INESC TEC and University of Minho
Porto, Portugal
georges.r.younes@inesctec.pt

ABSTRACT

The GDPR is proving to be on the hot topics when discussing data privacy within European boundaries. However, the new regulation imposes technical challenges, namely on the right to be forgotten (article 17). The ability to erase data on a generic database system ends-up being natural process and its application within the GDPR becomes the activation of a protocol. However, distributed ledgers, particularly blockchain, arise as a new way to store data, surrounding it with key properties for ensuring trust. In this paper we cover and envision how sensitive data can leverage on blockchain technologies, while allowing data to be promptly removed at the users discretion.

1 INTRODUCTION

The cloud computing paradigm together with the massive use of mobile technologies triggered a new standard on how people access data. However, the major investment has been done towards making such applications ubiquitous, enabling them to be accessible on-demand, and a minor investment was spent towards ensuring data privacy.

The recently introduced General Data Protection Regulation (GDPR) [4] tries to overcome data privacy concerns by empowering citizens with strict control of their data. Moreover, it forces entities that hold or process data to be explicit on what is achieved with the data and to acquire strict consent on that basis. It empowers a citizen to have control of data by revoking consents. An analysis of the GDPR guides us to argue that it was designed with traditional data persistence mechanisms in mind, namely centralized database systems.

However, decentralized ledger technologies, that is, Blockchain [1–3, 5] have currently gained a lot of attraction. They are based on a network of nodes (that do not necessarily trust each other) that jointly agree on the validity of data that is pushed by each one of the nodes. Periodically, new data changes are produced and as a result of an agreement by a majority of the peers, a new block is produced. As new blocks emerge, links to previous blocks are established; therefore creating a chain. Changing the data inside one of these blocks becomes infeasible as it would imply to reassemble all the links down the chain and across more than half the nodes in the network. Since new pieces of data are the result of an agreement derived from a majority of nodes, and since it is infeasible to modify blocks, Blockchains provide two important properties, that is, immutability and trust. Is this set of properties, often materialized in non-repudiation and the ability that all nodes have to validate data, that these technologies have an edge over single domain data management systems.

In light of the regulatory changes imposed by the GDPR, applications and business use-cases that depend on blockchains fall under a paradox regarding the execution of article 17, that is the Right To Be Forgotten (RTBF). The immutability and associated ability to validate data, the main differentiators and from where value can be drawn, are actually the ones that directly impose a threat when complying with GDPR and granting the ability to erase data.

Blockchain technologies along with all available instantiations display a set of very attractive features to applications. However, as such applications are required to comply with the GDPR, the best solution should be derived from both sides.

In this vision paper, we briefly cover distinct types of distributed ledger technologies on our way to provide an analysis on 3 distinct axis. We then explore one possibility to overcome the paradox, by proposing a mechanism approaching blockchain's trust insurance and a third-party system to store sensitive data off-chain.

2 THE PARADOX: GDPR VS BLOCKCHAIN

In a world where data is the most valuable asset, GDPR's main goal was to extend the protection of individuals right for privacy and reach the "virtual territories". Those rights include data ownership, mandatory consent, RTBF, etc; which puts the individual back in control of their own data. However, it seems that at the time the GDPR was being developed, it was meant to target conventional database systems and therefore was not focused on technologies such as the blockchain.

The attractiveness of blockchain stems from an unconventional and original way of allowing fully decentralized systems where trust is not needed. The blockchain brings many benefits but more importantly, it introduces new concepts such as immutability and transparency. The immutability of the blockchain guarantees that data stored, once validated, is tamper-proof and therefore provides integrity and non-repudiation. However, the immutability property prohibits the right to be forgotten, in article 17 of the GDPR. Transparency, as well, is a property of the blockchain (public blockchains) which guarantees for the users of the blockchain, that no party can mask or hide data. This is a very interesting property to have in public sector applications, where everything that happened is logged and public. However, choosing transparency is also forfeiting privacy, which is an important property to have in a system where data is shared, as well as a right protected by the GDPR.

The rise of the blockchain and blockchain-based applications and the increasing interest and adoption of this technology by individuals, businesses and governments can not be ignored, whether people believe it to be a disruptive technology or not. But, any eventual adoption and use of this technology, at least in the EU and

when the data subjects/processors/controllers are located in the EU, must comply to the GDPR rules.

3 PROPOSAL

Application's compliance with the GDPR falls into solving the paradox between the key factors when choosing blockchain to support or manage data, and the ability to ensure erasure capability. Likewise, our vision falls under a theorem like structure with 3 basic pillars, where we slotted the previously addressed constraints.

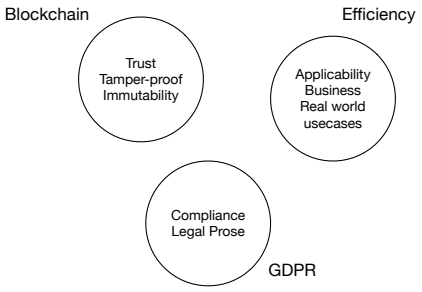

Figure 1: Properties universe.

Figure 1 depicts them. Pillar *BC* characterizes the blockchain side of the spectrum, representing requirements such as being immutable, tamper-proof and transparent, that is, that data is not modified after being committed and that changes are traceable and verifiable by anyone. Pillar *Efficiency* characterizes the need for efficiency that real-world business use cases require. Pillar *GDPR* characterizes the need to be compliant with legislation, not necessarily, but also including the GDPR.

The paradox is established on the basis that is not possible to cover all these three pillars. Likewise, we establish 3 distinct levels:

Public

This level is characterized by: $\overline{Efficiency} + \overline{GDPR} + BC(T + NR)$, *i.e.,* Trust + Non-Repudiation. Public ledgers provide the guarantees of public blockchain technologies (*i.e.,* BC), but the mechanisms in state that grant the ability to be tamper-proof disable an efficient execution, as this procedure extends through time. Moreover, as it is infeasible to change data in due time, it is not possible to provide *GDPR* compliance, particularly on the basis of article 17.

Single Domain

This level is characterized by: $Efficiency + GDPR + \overline{BC}(\overline{T} + NR)$. Single Domain solutions such as database systems can be adjusted to particular use cases scenarios. Therefore, Efficiency is granted by fine tuning the system to accommodate performance requirements. Even for the case of distributed configurations, the single domain install, trust is granted, not incurring in an efficiency penalty. Moreover, compliance with GDPR's right to be forgotten is possible as the underlying data model is mutable, allowing data deletion. However, as data is kept within the boundaries of a single entity, the tamper-proof property cannot be granted, as other parties can't validate the integrity of data.

Consortium

This level is characterized by: $Efficiency + GPDR + \overline{BC}(T + \overline{NR})$. Consortium based ledgers allow a mixture between of properties from public blockchains and single domain systems. The non-holistic view of data favours efficiency, as not all nodes take part in reaching agreement. Still, as there are usually sets of special nodes, that are globally trusted and ensure data validity, it is possible to ensure trust across all nodes. The compliance with GDPR (*i.e.,* allowing data to be erased) may be possible if meaningful data is not stored on the blockchain itself. That is, the blockchain can be used to track state changes and allow the integrity check of data that is stored on a third-party service.

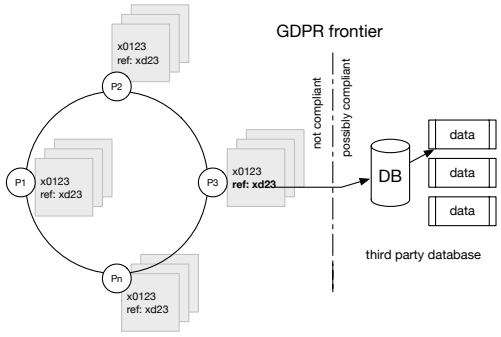

Figure 2: Architecture with on-ledger trust and off-ledger data.

Figure 2 depicts the vision in this paper. Data that would be enclosed under a GDPR consent would be placed in third-party database system, with its own access control mechanisms. As data is made durable in the database system, a digest is computed considering the data stored. Afterwards, a new transaction is assembled, considering the digest as the transaction data and pushing it to the ledger. The generated transaction ID is then used to identify the object in the database system. This architecture provides the ability for other nodes to verify a particular data item and by re-computing the digest and comparing it with the value on the blockchain; thus validating if it was tampered with. However, since data it kept in a third-party, only users with access to that system are able to make the verification.

This scheme ensures the right to be forgotten under the assumption of a single domain level. Data is able to deleted from the third-party system, breaking the reference that existed on the blockchain.

ACKNOWLEDGEMENTS

This work is financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project «POCI-01-0145-FEDER-006961», and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia as part of project «UID/EEA/50014/2013» and Project "TEC4Growth - Pervasive Intelligence, Enhancers and Proofs of Concept with Industrial Impact/NORTE-01-0145-FEDER-000020" is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

REFERENCES

[1] Richard Gendal Brown. 2018. The Corda Platform: An Introduction. (2018).
[2] Christian Cachin. 2016. Architecture of the hyperledger blockchain fabric. In *Workshop on Distributed Cryptocurrencies and Consensus Ledgers*, Vol. 310.
[3] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. (2008).
[4] Paul Voigt and Axel Von dem Bussche. 2017. *The EU General Data Protection Regulation (GDPR)*. Vol. 18. Springer.
[5] Gavin Wood. 2014. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper* 151 (2014), 1–32.