The Blockchain Role in Ethical Data Acquisition and Provisioning

Sara Migliorini[1,2], Mauro Gambini[1,2], Alberto Belussi[1], and Carlo Combi[1]

[1] Department of Computer Science, University of Verona, Italy
{name.surname}@univr.it
[2] Member of the IEEE Blockchain Technical Community

Abstract. The collection of personal data through mobile applications and IoT devices represents the core business of many corporations. From one hand, users are losing control about the property of their data and rarely are conscious about what they are sharing with whom; from the other hand, laws like the European General Data Protection Regulation try to bring data control and ownership back to users. In this paper we discuss the possible impact of the blockchain technology in building independent and resilient data management systems able to ensure data ownership and traceability. The use of this technology could play a major role in creating a transparent global market of aggregated personal data where voluntary acquisition is subject to clear rules and some forms of incentives, making not only the process ethical but also encouraging the sharing of high quality sensitive data.

Keywords: Blockchain · Decentralized Autonomous Organization · Network coalition · Data ownership and traceability · Voluntary provisioning.

1 Introduction

The acquisition and processing of personal data through mobile applications and IoT devices is the core business of many IT corporations. The amount of generated data is predicted to reach 44ZB by 2020, at the same time IoT devices will be around 30 billions [8]. Nowadays, individuals have very uncertain control about their data, how they are collected, viewed and monetized. In the future all major countries are likely to introduce specific privacy laws for protecting personal data. For example, any organization that provides goods and services to EU citizens must comply with the new General Data Protection Regulation (GDPR) [5]. The GDPR protects any information that can be directly or indirectly used to identify a person. This information varies from user names, emails and IP addresses to healthcare information and bank details. GDPR includes several rights, for instance the right to be forgotten (GDPR Art.17) and the right to be notified of data breaches (GDPR Art.33–34). Any data processing

PIE 2019, June 4, 2019, Rome, Italy. Copyright held by the author(s).

that is not compliant with the GDPR can result in significant fines, till Euros 20 million or 4% of the global company revenues. Laws like the GDPR increase the security requirements of any business that processes personal data and that means more burden and risks for the involved companies. Big IT corporations can manage the risk by delegating to new specialized companies the collection and aggregation of personal data.

Open networks and public ledgers can provide an alternative business model to control and trace the use of personal data. The blockchain can lead to the development of independent and resilient data management systems able to ensure data ownership and traceability and increasing the user awareness [12]. This may guarantee a more fair use of the data and it can be the only viable way to collect sensible data on voluntary base, like healthcare related information and biological profiles. Nevertheless, the adoption of a blockchain infrastructure comes with its own limits and the open issues are manifold. For instance, relatively to the right to be forgotten (GDPR Art.17), blockchain deletions, or more in general blockchain updates, should be carefully investigated [1, 9].

The aim of this paper is to take a look at the potentialities offered by this new technology in the development of new ways to collect, maintain and use personal data, and discuss some problems and limitations that have to be overcome for making it effective in this scenario. A blockchain infrastructure might be the right way to bring back to the users the ownership of their data. With a clear idea about what is shared and with whom, users can be encouraged to share more personal and sensitive data with specific companies, even on voluntary basis, revoking such privilege at their discretion. Moreover, this technology can also provide the right infrastructure for creating a global market of aggregated personal data: well-informed conscious users can decide to share their personal data on voluntary basis in presence of a clear usage rules and economic benefits.

The remainder of this paper is organized as follows: Sect. 2 summarizes some previous investigations about the applicability of the blockchain technology in the field of data provisioning. Sect. 3 briefly illustrates the main concepts underlying the blockchain technology and the notion of network coalition. Sec. 4 investigates the idea of developing a network coalition for voluntary data provisioning. Finally, Sect. 5 summarizes the paper and discusses future work.

2 Related Work

In recent years some blockchain-based solutions have been proposed for sharing medical data among several hospitals while providing data access control, provenance and auditing [13, 7]. However, blockchain was originally designed to record transactional data, which is relatively small in size, while the information to be stored can be large, for instance in the healthcare domain many images or treatment plans have to be recorded. The notion of off-chain storage has been proposed in literature to deal with this problem. Essentially, data are kept outside the blockchain, for instance in a traditional database, while the blockchain will only be used to store their digital fingerprints to ensure data authenticity [4].

Even if the blockchain was originally developed for storing only information about financial transactions, in these years it has also been used to certify the existence and tracking the ownership of digital or physical assets [3, 6]. It is estimated that Bitcoin transactions storing different information are about the 1% of the total transactions in the Bitcoin blockchain [1].

3 Blockchain and Network Coalitions

Currently, several variants of blockchain exist, these variants are often classified as *distributed ledgers*. In its original form [11] a *blockchain* is essentially a temporally ordered list of permanent data blocks. The head of the list is called *genesis block* and includes some evidence about its release date, while every other block is generated at fairly regular intervals and contains a cryptographic message digest, or briefly a *hash*, of its predecessor, creating a chain of references. Each block also includes a *proof of work*, namely an evidence that a certain amount of work have been spent for producing it. This proof is obtained by repeatedly applying a cryptographic hash function to a block, varying its content at each iteration by using a different nonce, until one of the target hashes is found. The described operation is part of the *mining* process that is simultaneously performed by several competitive network agents called *miners*. Altering a given block requires the recomputation of the hashes of all its successors in a limited amount of time. Since such operation could be quite expensive, the probability of observing a block replaced by another one decreases over time as new ones are added in front of it. A block referring to a given one is said to *confirm* it and after a certain number of confirmations, a block is considered practically immutable.

The key innovation of the blockchain technology is a decentralized emergent consensus protocol that enables a group of agents to reach an agreement about a global state by accepting data transmitted across an open byzantine Peer-to-Peer (P2P) network. The consensus can be considered *emergent*, because there is not specific point in time in which it is explicitly reached; while the network is said to be *open* and *byzantine*, because agents can be self-interested, they can enter and leave the system without authentication or secure connections and they can act strategically against the P2P protocol.

The blockchain technology appeared for the first time in the implementation of the Bitcoin protocol [11] as a clever solution to the double-spending problem that does not require a trusted central authority. In Bitcoin, each block contains a set of transactions representing the transferring of tokens from a source to a destination account address. Following the Bitcoin protocol, each agent can independently validate both transactions and blocks and reach a consensus about the blockchain state in an autonomous way.

A generalization of the Bitcoin protocol that properly extends the blockchain technology has been proposed by Ethereum [2] and similar platforms, in which a global state can be updated, not only by token transactions, but also by generic instructions previously stored in the blockchain. In particular, the Ethereum platform provides a virtual machine, called Ethereum VM or EVM, that can run general-purpose scripts encoding arbitrary state transition functions. These scripts are called *smart contracts* and they are considered autonomous software agents executed by the EVM when a certain event occurs, for instance when a transaction is scheduled or a message received. When a contract is triggered, it runs a sequence of predefined instructions that can control the related token balance, the key-value store used to keep track of persistent variables and the invocation of other contracts [2].

Ethereum contracts are sufficiently expressive to create new cryptocurrencies like Bitcoin and to found network coalitions [10], often called Decentralized Autonomous Organizations (DAOs). A network coalition is a concerted form of cooperation, in which a group of actors decide to collaborate with the explicit purpose to achieve a common goal. Supply chains, cooperatives, strategic business alliances, joint ventures can be good examples of coalitions. Decentralization, namely the lack of an established central authority, is a main characteristic, together with the possibility to have a dynamic composition, that is new components can freely join the coalition, while existing ones can leave it. Governance rules are encoded inside a set of smart contracts and members hold a certain amount of tokens through which they can exercise their voting power.

4 Voluntary-based Coalitions for Data Provisioning

A blockchain infrastructure might be the right way to bring back to the users the ownership of their data. With a clear idea about what is shared and with whom, users can be encouraged to share more personal data with specific companies, even on voluntary basis, revoking such privilege at their discretion. Moreover, this technology can also provide the right infrastructure for creating a global market of aggregated personal data: well-informed conscious users can decide to share their personal data on voluntary basis in presence of a clear benefit. However, the utility of a single piece of personal data is difficult to quantify and its value is typically very low. The value of personal data can increase only after some aggregation and integration process, such process is usually performed by external companies which collect data from users and resell the aggregated

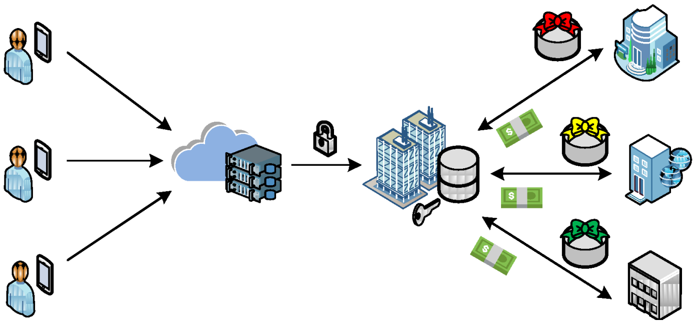

Fig. 1. Representation of the current process in user data provisioning.

datasets to other companies. The current provisioning process is depicted in Fig. 1, where users share in a more or less conscious way their personal data to an organization, depicted in the middle, which takes care of aggregating such raw data and producing useful information that in turn will be sold to other organizations, depicted on the right. The users are unaware of how such data are processed by the middle organization and have no control about the nature of the organizations on the right and the usage they can make of these data.

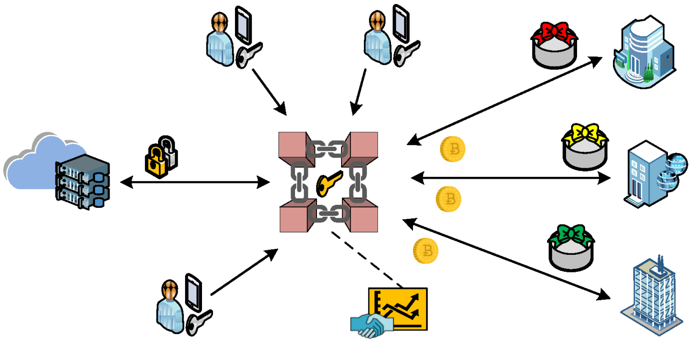

Fig. 2. Innovative data provisioning paradigm induced by the use of the blockchain technology and the establishment of a voluntary-based DAO.

A blockchain infrastructure can trigger a paradigm shift in the acquisition and aggregation process. With the right technology, users can voluntary collaborate for collecting and aggregating their personal data, producing valuable information. The coalition can sell such data to other organizations, but potentially maintaining the control about their use and transfer. This new form of data collection and aggregation is exemplified in Fig. 2 where the centralized company in Fig. 1 is replaced by a network coalition formed by an open P2P network of users. This paradigm shift can be a win/win condition for both users and companies: from one hand, users have more control on their personal data and they are more conscious of their roles and rights as a group. For instance, they can move a class action against improper data usage when this cannot be prevented with cryptographic methods. The blockchain can act as a tamper-proof log and used as evidence before the court in case of dispute. From the other hand, companies can externalize some data acquisition costs and reduce the risk induced by a wrong treatment of personal data w.r.t. the existing privacy regulations. In addition, this new form of data acquisition can be an effective way to collect huge amount of personal data that require a voluntary and incentivized effort.

5 Conclusion and Future Work

This paper takes a first look to the applicability of the emerging blockchain technology for building independent and resilient data management systems able to ensure data ownership and traceability. The blockchain technology is considered an enabling technology, namely a technology that opens the design space to new innovative applications and even to new way of thinking about algorithmic solutions in which economic aspects play a major role. We conjecture that this technology may be useful to both encourage users to share their data even in more sensitive context, such as the health-care one, and to create a global market of aggregated personal data. Despite the benefits of using a blockchain infrastructure for data provisioning, its adoption comes with its own limits and the open issues are manifold. For instance, relatively to the right to be forgotten (GDPR Art.17), blockchain deletions, or more in general blockchain updates, should be carefully investigated [1, 9]. Relatively to the actual establishment of network coalitions for data provisioning, particular attention has to be placed to their legal recognition in different countries and to the way privacy laws, like the GDPR, can be applied to them.

References

1. Bartoletti, M., Pompianu, L.: An Analysis of Bitcoin OP_RETURN Metadata. In: Financial Cryptography and Data Security. pp. 218–230 (2017)
2. Buterin, V.: A Next-generation Smart Contract and Decentralized Application Platform (2014), `http://github.com/ethereum/wiki/wiki/White-Paper`
3. Chen, J., Xue, Y.: Bootstrapping a blockchain based ecosystem for big data exchange. In: 2017 IEEE International Congress on Big Data. pp. 460–463 (2017)
4. Esposito, C., De Santis, A., Tortora, G., Chang, H., Choo, K.R.: Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy? IEEE Cloud Computing **5**(1), 31–37 (2018)
5. EU Commission: General Data Protection Regulation (GDPR), Regulation EU 2016/769., `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504`, (acc.: 2019-03)
6. Karafiloski, E., Mishev, A.: Blockchain solutions for big data challenges: A literature review. In: 17th Int. Conf. on Smart Technologies IEEE. pp. 763–768 (2017)
7. Kim, H.E., Kuo, T.T., Ohno-Machado, L.: Blockchain distributed ledger technologies for biomedical and health care applications. Journal of the American Medical Informatics Association **24**(6), 1211–1220 (09 2017)
8. Kugler, L.: The War over the Value of Personal Data. Communications of the ACM **61**(2), 17–19 (2018)
9. Matzutt, R., Henze, M., Ziegeldorf, J.H., Hiller, J., Wehrle, K.: Thwarting Unwanted Blockchain Content Insertion. In: 2018 IEEE International Conference on Cloud Engineering (IC2E). pp. 364–370 (2018)
10. Migliorini, S., Gambini, M., Combi, C., La Rosa, M.: The Rise of Enforceable Business Processes from the Hashes of Blockchain-Based Smart Contracts. In: Enterprise, Business-Process and Information Systems Modeling. pp. 130–138 (2019)
11. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008), `http://www.bitcoin.org/bitcoin.pdf`, (acc.: 2018-11)
12. Shafagh, H., Burkhalter, L., Hithnawi, A., Duquennoy, S.: Towards blockchain-based auditable storage and sharing of iot data. In: Proceedings of the 2017 on Cloud Computing Security Workshop. pp. 45–50. CCSW '17 (2017)
13. Xia, Q., Sifah, E.B., Asamoah, K.O., Gao, J., Du, X., Guizani, M.: MeDShare: Trust-Less Medical Data Sharing Among Cloud Service Providers via Blockchain. IEEE Access **5**, 14757–14767 (2017)