Using blockchain to decentralize and protect user privacy in compliance with GDPR

Stefan Dejanovic*, Jelena Marjanovic*, Imre Lendak*, Alekdandar Erdeljan*
University of Novi Sad, Faculty of Technical Sciences
Novi Sad, Serbia
{stefan.dejanovic, jelena.stankovski, Lendak, ftn_erdeljan}@uns.ac.rs

Abstract—

In this paper, we will analyze how can we use blockchain technology for access control, that does not require trust in a third party. Current model, in which third parties collect and control massive amounts of personal data is questioned, because there are many reported incidents of security breaches that are compromising user's personal data. On the other end, we have new EU directive, General Data Protection Regulation (GDPR), that is aiming to achieve complete protection of personal data in the EU and for the free movement and removal of such data. We describe personal data management system that is decentralized that ensures users are the owners and controllers of their data. Finally, we analyze how solution is GDPR compliant and possible future extensions.

I. Introduction

One of the current hot research topics, both in industry and academia, is blockchain. Defined as a shared distributed ledger, blockchain uses a set of nodes to maintain data structure, organized in blocks.

We are in the Big Data era, data is constantly being collected and analyzed. Organizations are using the data they are collecting to improve their service and optimize their process and decisions. Now, data is one of the most valuable assets in economy [5]. On the other end, while we are data-driven society and constantly creating and storing more data, one of the growing public concerns is user privacy. Centralized organizations are storing large amounts of personal and sensitive information. Individuals have no or little control over their personal data that is stored and how it is used. The latest legislation created for the EU, GDPR, aims to achieve complete protection of user's personal data. At the time GDPR was created, we only used centralized cloud services and data collection business models that persist as the main source of revenue for companies. From then, decentralized systems started developing rapidly, and may require adjustments to the GDPR directive. Motivation in this paper stands in perceiving main benefits of using blockchain technology as access control manager that does not require third party trust and complying with GDPR directive. The remainder of this paper is organized as follows: Section II reviews related works regarding blockchain technology and protecting personal data. Section III describes methods used in this analysis. Section IV presents results obtained in this analysis. Section V summarizes our testing and gives a scope for future work.

II. Related Works

Vast number of industries have shown interest in blockchain technologies and these privacy issues, both from a legislative point of view and from a technological point of view [6]. Leading examples of companies and organizations chose to implement their own proprietary authentication methods based on the OAuth protocol [4].

There are various solutions and techniques targeting privacy concerns that are focusing personal data. Data anonymization methods are aiming to protect personal and sensitive data. K-anonymity is a common property of anonymized datasets and it requires that sensitive information of each record is indistinguishable from at least k-1 different records [3]. Furthermore, there is l-diversity that is used to ensure that sensitive data is represented by a diverse enough set of possible values [2]. Also, t-closeness is used for distribution of sensitive data [1].

There is a research [7] that has demonstrated how anonymized datasets employing these techniques can be de-anonymized. Other privacy techniques and implementations are based on differential privacy, a solution that perturbs data or adds noise to the computational process before sharing personal and sensitive data [8].

Lastly, in recent years, new technology emerged which allows users to transfer currency (crypto) securely without a centralized authority, that uses publicly verifiable open ledger. Example of that system is Bitcoin. From then, lot of projects demonstrated how blockchain technology can be used to serve other functions requiring trusted computing.

A. Our Work

We are using combination of blockchain technology and off-chain storage to build a personal data management platform with focus on maintaining user privacy. We will describe our solution and discuss future improvements, how blockchain technology can be used as a core resource in trusted computing.

III. METHODS

In this section methodology that is used to evaluate the blockchain platforms is going to be described.

A. Research Questions

In order to analyze main benefits of using blockchain technology as access control manager and complying that model with GDPR, certain concepts must be comprehended.

First one is data ownership. Our model focuses on ensuring that personal data is owned by users. System recognizes users as the owners of the data and the service applications as guests with permissions.

Second is data transparency and auditability. Every user has complete transparency over data that is being collected about him and how that data is being accessed.

Third is access control. Users are required to grant set of permissions to the service applications. Opt-in when application is requesting permission to access user's personal data and opt-out when permissions are altered.

B. Methodology

We will rely on the blockchain to be tamper-free and no misuse can be made, because it requires large network of untrusted peers to compromise blockchain security. Also, we assume that user manages his keys securely, using secure wallet service. We will concentrate on analyzing inputs and outputs of transactions where data is transferred with certain permissions. We will analyze what is written on the blockchain and data storage and analyze if data is encrypted (hashed). Test scenarios will be created where input and output is going to be measured and analyzed are:

1. Application is issuing transaction to fetch user's data with permission
2. Application is issuing transaction to fetch user's data without permission
3. User revoked permission for service application
4. User alters his personal data
5. User deletes his personal data

C. Solution

System is illustrated in Figure 1. It is consisted of three entities:

1. Users, who add and use service applications;
2. Services, providers of such service applications that are using user's personal data
3. Nodes, entities in the blockchain that are maintaining data structure and a auxiliary structure – distributed hash table.

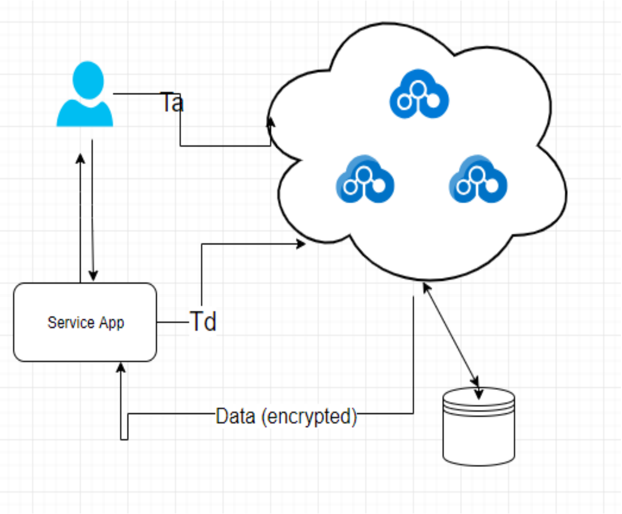

Figure 1. – Solution design

The Blockchain part of the system accepts two transactions:

- Ta – used for access control management and
- Td – used for storing and fetching data.

Example of the process:

User subscribes or registers for a service, Ta will be generated and sent to the blockchain. Ta transactions is consisted of identity pair (user, service) and associated permissions. That will represent permissions for that service for using user's personal data. Data that is collected from the user side is encrypted using a shared encryption key and sent to the blockchain as a Td transaction. Data is also sent to the distributed hash table (off-chain) with Td transaction. Service application will only retain a pointer to the data on the blockchain. Pointer is hash of the data.

Now both user and service can query user's personal data using Td transaction with pointer. When Td is issued, blockchain will verify digital signature of user or service that is accessing the data. If service is accessing the data, permissions will be checked as well. User can issue new Ta transaction with new permission set and change or revoke permission for service. Dashboard is developed to allow and show overview of the user's data, permissions and where user can manage his permissions. Dashboard will be like crypto wallets such as for Ethereum and Bitcoin. On the dashboard we will track how data is processed, changed and deleted, how permissions are granted and revoked to analyze the model and compliance with GDPR. Dashboard will be small decentralized application, with connection to blockchain network. For this purpose, Hyperledger Fabric blockchain is used and dashboard is connected with Hyperledger Fabric blockchain using RESTful API.

IV. RESULTS

A. Privacy Analysis

We measured results by test scenarios:

1. When service is accessing user data with permission, service receives encrypted data and transaction gets recorded on the blockchain, user can see transaction on the dashboard log.
2. When application is accessing user data without permission, service receives information that it does not have permission to access that data, transaction is recorded and showed on the user dashboard log.
3. When user revokes permission for service, transaction gets recorded and new Ta transaction is created for user, service pair and sent to the blockchain.

B. GDPR Analysis

We analyze GDOR compliance with following test cases:

1. Right to rectification (Art. 16) – User changes his personal data
   a. New data is created, encrypted and stored on the blockchain, with new hash value (pointer for the data) with new shared key
   b. Old data is on the blockchain, but encryption key is deleted from DHT, data on the blockchain is no more accessible
2. Right to erasure (Art. 17) – User deletes his personal data
   a. Ta is issued with revoking all permissions from all services that have permission for user's personal data
   b. Data is on the blockchain, but encryption key is deleted from DHT, data on the blockchain is no more accessible
   c.

Model that is designed this way, ensures only user has control over his data. Blockchain as decentralized technology combined with signed transactions ensure that services or intruders cannot pose as the user, or corrupt the network, because intruder needs to own user's digital signature or own more than half of the network. Intruder cannot learn anything from the blockchain because data is encrypted that none of the nodes has.

V. DISCUSSION

Personal and sensitive data should not be trusted with third party. Third party can be susceptible to attacks and misuse. Users should own and control their personal data without compromising security. Our model provides that solution, using blockchain as access control manager and off-chain distributed storage. Key benefit is that blockchain recognizes users as the owners of their personal data. Users can be data controllers and service providers can focus on utilizing the data, rather than controlling and securing personal data. With this model, legal and regulatory decisions about GDPR should be simpler, even regulations can be implemented into blockchain so they can execute automatically.

Issue with GDPR and Art16 & 17 is because GDPR was created when centralized systems are mainly used. Even deleting shared encryption key, data remains on the blockchain. That can be interpreted ambiguously by regulators. Regulation needs to follow technology to avoid such miss interpretations.

REFERENCES

[1] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In ICDE, volume 7, pages 106–115, 2007

[2] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity.ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1):3, 2007.

[3] Latanya Sweeney. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05):557–570, 2002.

[4] Juan Perez. Facebook, google launch data portability programs to all, 2008.

[5] K Schwab, A Marcus, JO Oyola, W Hoffman, and M Luzi. Personal data: The emergence of a new asset class. In An Initiative of the World Economic Forum, 2011.

[6] Rt.com. Obama announces legislation protecting personal data, student digital privacy, 2015.

[7] Arvind Narayanan and Vitaly Shmatikov. How to break anonymity of the netflix prize dataset. arXiv preprint cs/0610105, 2006.

[8] Cynthia Dwork. Differential privacy. In Automata, languages and programming, pages 1–12. Springer, 2006.