Blockchain-Based Data Sharing and Managing Sensitive Data

Ahmet Fatih Mustaçoğlu*  
1 Bilgi Güvenliği Mühendisliği, İstanbul Şehir Üniversitesi, İstanbul, Turkey

(First received 3 July 2018 and in final form 26 November 2018)  
(DOI: 10.31590/ejosat.440049)

Abstract

IT advancements have produced huge amount of data including personal and sensitive information. People have no control on data that is stored, processed and controlled by third parties who could harm user privacy. Meanwhile blockchain technology has potential to drive research and applications and will weave together as we look at digital economy processing personal data. In this research work, a blockchain that is underlying technology behind Bitcoin based access control mechanism is introduced to protect users’ privacy. A blockchain is a growing list of distributed records that are connected to each other through the usage of cryptography. The distributed consensus and anonymity form two major characteristics of the blockchain technology. The proposed system implements a protocol that utilizes blockchain technology to manage access control to users’ data without depending on a third party.

Key words: Blockchain, Bitcoin, privacy, anonymity.

1. Introduction

Amount of data in today’s eco system is exponentially increasing. Based on a recent report [1], twenty percent of the world’s data has been accumulated during the past couple of years. The social media tools have the biggest contribution to this occurrence. Facebook that is the most preferred online social-network collected 300 petabytes of personal data since its existence [2]. Today’s world, data is critical for every business and it is regularly being retrieved and analyzed resulting in technological advancements and economic growth for countries. Companies and institutions benefit from the collected data in order to optimize their costs and their critical processes in order to compete with others and more. So, data is an important asset in every business on the world and there exist many benefits of data-driven society [3].

The user privacy is a growing public concern that need to be considered carefully in today’s data-driven society. Centralized institutions or companies process large amount of personal and sensitive data that are both can be public or private. People have little or no knowledge about the data that is stored about them and how it is used. In recent years many incidents related to privacy are published on public media. For example, the government surveillance [4] and a large-scale scientific experiment that was carried on without informing the participants [5].

Blockchain is a peer-to-peer (P2P) distributed ledger technology and it is the underlying structure for Bitcoin. It consists of three main components:

- A distributed network: Blockchain forms a decentralized P2P architecture with nodes composed of network members. Each node keeps an identical copy of the blockchain and plays a role for the process of validating and confirming digital transactions for the network.

- A shared ledger: It holds all the transactions. Nodes in the distributed networks record digital transactions into a shared ledger. Each member in the network executes algorithms to evaluate and verify the requested transactions in order to add the transactions. If the transaction is validated by the majority number of member in the network, then the new transaction is added to the shared ledger. Changes to a shared ledger are updated in all copies of the blockchain in couple of minutes.

- Digital transactions: A blockchain can hold any type of data or digital asset. The network that implements the blockchain determine the type of the information carried in the transaction. To guarantee authenticity and accuracy, information is also encrypted and digitally signed.

Blockchain technology is one of the best advancements since the internet itself. It provides a way to value exchange among people without the trust or central authority. Its major advantages are:

- There is no third party that carries the value and control over it
- The cost of a transaction from and to anywhere is cheap
- Any amount of value can be transferred in minutes and the transaction is considered as secure

---

*Corresponding Author: Bilgi Güvenliği Mühendisliği, Istanbul Şehir Üniversitesi, Istanbul, Turkey, ahmetfatihmustacoglu@sehir.edu.tr
- All the transactions are transparent so they can be verified at any time

- It provides a base for decentralized applications that would be able to transfer value and manage data.

**Related Work.** There exists many efforts in order to address these privacy concerns from technological and legal perspective [6, 7]. The patterns of personal information leakage and privacy implications associated with online networking are researched and evaluated in [8]. The key consumer perceptions of privacy is investigated and analyzed through consumer responses [9]. Many well-known corporates prefer to execute their in-house developed authentication software that is based on the OAuth protocol [10]. Then, they operate as a centralized trusted authority.

Along with the technological advancements, researchers have focused on privacy issues for personal data and many approaches have been proposed for the solution. Data anonymization is one of the privacy-preserving methods and it aims to hide personally identifiable information in either encrypting or removing from data. k-anonymity approach proposes a formal model for protecting privacy and a set of accompanying policies for deployment [11]. l-diversity and t-closeness are the related extensions to k-anonymity. The sensitive data is characterized by a heterogeneous enough set of possible values in l-diversity [12]. T-closeness focuses on the distribution of sensitive data [13]. Beyond these anonymization techniques, there also exists some approaches to break them [14, 15]. Encryption schemes and differential privacy are also major privacy-preserving techniques. Encryption schemes provide a mechanism to query and calculations over encrypted data like fully homomorphic encryption (FHE) [16]. Differential privacy guarantees privacy by making data ambiguous or adding noise to it.

**2. Users’ Data and the Privacy Problem**

The privacy issues that users face with regarding to their data coming from various online services are pointed out through the paper. Especially in cloud where users have no control how their data is stored and shared with other users. This paper primarily focuses on the privacy of users’ data that is stored and shared with other users through the usage of a blockchain as an access-control manager. In the presented approach, the underlying protocols are described in detail and the proposed approach protects against the following well-known privacy issues:

- **Ownership of Meta-Data.** The proposed system guarantees that users’ data coming from social media tools is own and control by the owner. So, the access permissions can be defined for a digital data by the owner for other users.

- **Transparency of Users’ Meta-Data.** Each user of the proposed system is aware of what digital data he has in the system and how the data shared and accessed by other users.

- **Well-define access control.** In the proposed system, data for each user is collected from various internet sources (e.g. Facebook, twitter, delicious, google scholar, etc.) automatically or manually, and the owners define the access rights for their metadata in order to define how it is accessed by other users. This is usually done by storing an access control list in a database that is located on a local host or a cloud. However, in the proposed model, the access control policies (policy files for defining users’ Access to data objects) will be securely kept on a blockchain in a decentralized fashion and the system is not depend on a third party.

**3. Proposed Approach**

The proposed system consists of data owners, other users, and nodes presented in Figure 1. The system provides its users with ability to download, collect and share digital data coming from various social sites. The download service allows users to import their metadata from social web tools into the proposed system. The collecting service uses integrated search engines (e.g. google, google scholar etc.) to search the internet for a given key words and then import the selected results in to the proposed system. The sharing service enables users to share their data with other users.
based on blockchain technology. The nodes are charged with maintaining the blockchain and a distributed data store that holds a private key-value pairs.

4. The Protocol

This section describes the underlying protocol that is based on the Bitcoin [23] of the proposed system in detail. The standard cryptographic structures is used in the proposed system: a) a symmetric encryption scheme specified by the 3-element (Egen: the generator, Eenc: an encryption algorithm, Edec: a decryption algorithm); b) a digital signature scheme defined by 3-element (Sgen: the generator, Ssig: a signature algorithm, Sver: a verification algorithm) implemented by using ECDSA with secp256k1 curve [23]; and a cryptographic function (H) for hashing operations implemented as an instance of a SHA-256 [24].

4.1. Preliminaries

In this section, the preliminaries used in the proposed blockchain-based data sharing system are defined formally. The underlying mechanism requires familiarity with Bitcoin [25] and blockchains.

4.1.1 User Identities

Blockchains support anonymous identity since every user can generate as many public-key as the user desires in order to increase the privacy. The proposed work uses a shared identity that contains an owner and a user credentials. The shared identity consists of key pairs for an owner and a user as well as a symmetric key that is used for encrypting or decrypting the data. The implementation of the shared identity is given in Algorithm 1. The publicly shared identity and the full identity including private key pairs and a symmetric key can be defined by 2-element and 5-element respectively:

$$\text{shared}_\text{identity}^{\text{public}}_{\text{owner, user}} = (\text{pub}_\text{key}^{\text{owner, user}}, \text{pub}_\text{key}^{\text{user, owner}})$$

$$\text{shared}_\text{identity}_{\text{owner, user}} = (\text{pub}_\text{key}^{\text{owner, user}}, \text{priv}_\text{key}^{\text{owner, user}}, \text{pub}_\text{key}^{\text{user, owner}}, \text{priv}_\text{key}^{\text{user, owner}}, \text{sym}_\text{key}^{\text{enc}})$$

```
function SharedIdentity (owner, user)
  // Owner and user builds a secure channel to communicate
  Owner executes:
  (pub_key_{owner, user}, priv_key_{owner, user}) \rightarrow E_{gen}()
  sym_key_{enc} \leftarrow E_{gen}()
  Owner shares pub_key_{owner, user} and sym_key_{owner, user} with other user
```

In the proposed work, the off-blockchain key-value storage is implemented as an instance of Kademia [22]. Kademia is a distributed hashable (DHT) with added persistence. A network of nodes constituting the blockchain maintains the DHT, which is off the blockchain. DHT is also responsible for executing approved read and write transactions. Digital data is randomized across the nodes and replicated in order to provide high availability.

When the user interacts with the service (e.g. download or collect) to import metadata into the proposed system, a public or a private profile is generated and sent with the associated permissions to the blockchain in a Transaccess transaction. The digital data downloaded from a web tool by the regarding service is encrypted using a shared encryption key and sent to the blockchain in a Transdata transaction. Then the data is forwarded to an off-blockchain key-value storage and only a pointer to the digital data is left on the public ledger. The pointer is constructed from the SHA-256 hash of the data. Eventually, the digital data can be queried at any time on the blockchain through the usage of Transdata transaction with the pointer associated to it. First, the blockchain verifies the digital signature whether belongs to the owner or a non-owner user. Next, the access rights for the requesting user are retrieved on the blockchain and controlled. Then, the user granted access to the data if the appropriate access rights are exist. Moreover, the permissions for a data object can be updated at any time by the owner by invoking a Transaccess transaction with a new permissions.

In this section, the preliminaries used in the proposed system executes two type of transactions: Transaccess and Transdata. Transaccess is used for managing access controls and Transdata is used for handling digital data processed in related operations (e.g. store, download etc.). As an example; having logged into the system, a user downloads metadata from a social web tool into a database, which could be a local or a cloud, by using the proposed system for preserving his/her privacy.
User who needs an access to the data executes:

\[
(pub\_key^{user\_owner}_{sign}, priv\_key^{user\_owner}_{sign}) \leftarrow S_{gen}()
\]

//Now, both owner and user have: //sym\_key^{owner\_user}_{enc}, pub\_key^{owner\_user}_{sign} and pub\_key^{user\_owner}_{sign}

User shares pub\_key^{user\_owner}_{sign} with owner

Return sym\_key^{owner\_user}_{enc}, pub\_key^{owner\_user}_{sign}, pub\_key^{user\_owner}_{sign}

end function

Algorithm 1. Shared Identity Generation

4.1.2 Structure of the Blockchain

The blockchain is a distributed database that contains sequence of timestamped transactions. In the proposed work, let M be the blockchain memory space, represented as the hashtable

\[
M : \{0,1\}^{256} \rightarrow \{0,1\}^N
\]

where \(N \gg 256\). Transactions are stored and access through the hashtable M. Hashable key is 256-bit hash of a public key. When looking up a Trans\_access for an owner “k”, we will be executing \(M[H(public\_key_k)]\), however for looking up a Trans\_data, we will be executing \(M[H(public\_key_k||document\_id)]\). Trans\_access transaction has an output address for the associated Trans\_data transaction and Trans\_data transaction has an output address for the offline distributed hashtable for the encrypted data.

4.1.3 Data Access Policy

A data access policy contains a list of data ids that are granted access by an owner for a user, and they are accessible by the user. It is represented by POLICY\_{owner,user}. For example, if a user requested to access to object-1, object-5, object65…object-n, then

\[
\text{POLICY\_{owner,user}} = \{\text{object\_id-1, object\_id-5, object\_id-65…object\_id-n}\}
\]

4.1.4 Helper Functions

A message that is sent to a transaction needs to be parsed and Parse(x) function is used for this purposes. It deserializes the message. Furthermore, CheckPermissions\((pub\_key^k_{sign}, object\_id_x)\) function given in Algorithm 2 is used for checking a user’s permissions from the POLICY file located on the blockchain.

function CheckPermissions\((pub\_key^k_{sign}, object\_id_x)\)

\[r \leftarrow 0\]

\[\text{key\_policy} = H(pub\_key^k_{sign})\]

if \(M[\text{key\_policy}] \neq 0\) then

if \(pub\_key^{owner\_user}_{sign} = pub\_key^{user\_owner}_{sign}\) or \(pub\_key^k_{sign} = pub\_key^{owner\_user}_{sign}\)

\((pub\_key^k_{sign} = pub\_key^{user\_owner}_{sign}\text{ and object\_id}_x \in\text{POLICY\_{owner,user}})\) then

\[r \leftarrow 1\]

end if

end if

return r

end function

Algorithm 2. Checking a User’s Permissions from the blockchain

4.2. Protocols executed on the Blockchain

In this section, the major protocols that are performed on the blockchain when a Trans\_data or a Trans\_access transaction is requested are described in detail. The Access control protocol is carried out when a Trans\_access transaction is received and the details of the protocol is given in Algorithm 3. In order to read or write the encrypted data from/to the distributed hash table, the proposed system enforces the given protocol explained in detail in Algorithm 4.
function PerformAccessTransaction(pub_key^k_{sign}, transaction)
    r ← 0
    pub_key^k_{owner.user}, pub_key^k_{user.owner}, POLICY_{owner.user} ← Parse(transaction)
    // if it is owner
    if pub_key^k_{sign} = pub_key^k_{owner.user} then
        M[H(pub_key^k_{sign})] = transaction
        r ← 1
    end if
    return r
end function

Algorithm 3. The Access Control Protocol

function PerformDataTransaction(pub_key^k_{sign}, transaction)
    data, object-id, read_write = Parse (transaction)
    if CheckPermissions(pub_key^k_{sign}, object-id) = true then
        pub_key^k_{owner.user}, pub_key^k_{user.owner}, POLICY_{owner.user} ← Parse(M[H(pub_key^k_{sign})]) //Parse Transaction
        key_object-id = H(pub_key^k_{owner.user}) [object – id] //key for Transaction
        if read_write = 1 then // for writing read_write=1, for reading read_write=0
            data_hash = H(data)
            M[key_object-id] + data_hash
            (DHT) dht[data_hash] ← data //encrypted data
            return dht[data_hash]
        else if data_hash ∈ M[key_object-id] then // read permission only
            (DHT) return dht[data_hash] // return encrypted data for read
        end if
    end if
    return 0
end function

Algorithm 4. The Distributed Hash Table and Read/Write Protocol

4.3. Privacy and Security Analysis of the Proposed Work

In today’s digital world, security of sensitive data and privacy is a crucial topic for everyone. When keeping or sharing data with a third-party there exists a trust issue. Since, data could be exposed to attackers or could be misused as well. Hence, users should own and be in charge of managing their data without compromising security. The proposed work takes cares of the privacy of users’ data that is stored and shared with other users through the usage of a blockchain as an access-control manager blockchain technology and digitally-signed transactions that are the underlying mechanism of the Bitcoin.

Furthermore, the attacker cannot learn anything from the public ledger that are held at the distributed nodes of the blockchain network, since only hashed data are stored on the nodes. The data integrity risk can also be minimized through the enough distribution and the replication of data.

If an attacker is able to control one or more DHT nodes, the attacker still cannot learn anything about the data since it is encrypted with keys that are not kept on nodes.

Finally, a shared identity that is generated between an owner and a non-owner user guarantees that only a small part of data is obtained by an attacker in the event of an attacker possesses both the signing and encryption keys. If an attacker possesses just one of the keys then the data will be still secure.

5. Conclusion

In today’s digital world, security of sensitive data and privacy is a crucial topic for everyone. When keeping or sharing data with a third-party there exists a trust issue. Since, data could be exposed to attackers or could be misused as well. Hence, users should own and be in charge of managing their data without compromising security. The proposed work takes cares of the privacy of users’ data that is stored and shared with other users through the usage of a blockchain as an access-control manager.
in the blockchain. As the development of blockchain technology is still at an early stage, we hope our work will provide a better understanding of the design challenges of blockchain technology, and pave the way for further research in this area.

References


