



**Daffodil**  
*International*  
**University**

## **Towards A Blockchain Based Platform to Share the Datasets**

By

**D K Tonoy kumar**  
**(151-35-863)**

And

**Takia Islam**  
**(151-35-1014)**

A thesis submitted in partial fulfillment of the requirement for the degree  
of Bachelor of Science in Software Engineering

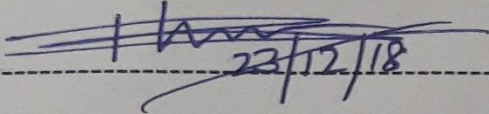
**Department of Software Engineering**  
**DAFFODIL INTERNATIONAL UNIVERSITY**

Fall-2018

## APPROVAL

This Thesis titled "Towards A Blockchain Based Platform to Share the Datasets", submitted by D K Tonoy kumar and Takia Islam, ID: 151-35-863 and ID: 151-35-1014 to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc in Software Engineering and approved as to its style and contents.

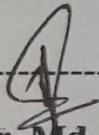
## BOARD OF EXAMINERS



23/12/18

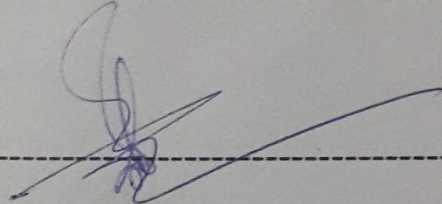
**Dr. Touhid Bhuiyan**  
**Professor and Head**  
Department of Software Engineering  
Faculty of Science and Information Technology  
Daffodil International University

**Chairman**



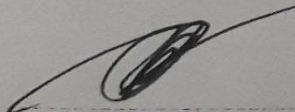
**Dr. Md. Asraf Ali**  
**Associate Professor**  
Department of Software Engineering  
Faculty of Science and Information Technology  
Daffodil International University

**Internal Examiner 1**



**Md. Maruf Hassan**  
**Assistant Professor**  
Department of Software Engineering  
Faculty of Science and Information Technology  
Daffodil International University

**Internal Examiner 2**



**Prof Dr. Mohammad Abul Kashem**  
**Professor**  
Department of Computer Science and Engineering  
Faculty of Electrical and Electronic Engineering  
Dhaka University of Engineering & Technology, Gazipur

**External Examiner**

## DECLARATION

We hereby declare that we have taken this thesis under the supervision of **SHEIKH SHAH MOHAMMAD MOTIUR RAHMAN**, Lecturer, Department of Software Engineering, Daffodil International University. We also declare that neither this thesis/project nor any part of this has been submitted elsewhere for award of any degree.

*Tonoy Kumar.*

---

**D K Tonoy Kumar**  
ID: 151-35-863  
Batch : 16<sup>th</sup>  
Department of Software Engineering  
Faculty of Science & Information Technology  
Daffodil International University

*Takia Islam*

---

**Takia Islam**  
ID: 151-35-1014  
Batch : 16<sup>th</sup>  
Department of Software Engineering  
Faculty of Science & Information Technology  
Daffodil International University

Certified by:

*AKR*  
*24.12.18*

---

**SHEIKH SHAH MOHAMMAD MOTIUR RAHMAN**

**Lecturer**

Department of Software Engineering  
Faculty of Science & Information Technology  
Daffodil International University

## **ACKNOWLEDGEMENT**

This work was guided by my reverent instructor SHEIKH SHAH MOHAMMAD MOTIUR RAHMAN. We thank our group member from Daffodil International University who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

We thank Takia for assistance with particular technique, methodology, Mrinmoy Sarkar for comments that greatly improved the manuscript.

We would also like to show our gratitude to the Kaushik sarker, Anwar Hossain for sharing their pearls of wisdom with us during the course of this research.

## **Dedication**

**Every challenging work needs self-efforts as well as guidance of elders especially those who were very close to our heart.**

**My humble effort I dedicated to my sweet and loving**

## **Father and Mother**

**Whose affection, love, encouragement and prays of day and night makes me able to get such success and honor.**

**Along with all hard working and respectful**

## **Teachers**

## TABLE OF CONTANT

### Contents

<b>APPROVAL .....</b>	<b>ii</b>
<b>DECLARATION .....</b>	<b>iii</b>
<b>ACKNOWLEDGEMENT .....</b>	<b>iv</b>
<b>Dedication .....</b>	<b>v</b>
<b>TABLE OF CONTANT .....</b>	<b>vi</b>
<b>LIST OF FIGURE .....</b>	<b>ix</b>
<b>ABSTRACT.....</b>	<b>x</b>
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
1.1 Background.....	1
1.2 Motivation of the Research.....	2
1.3 Problem Statement.....	3
1.3.1 Trust on central authority .....	3
1.3.2 Potential single point of failure.....	3
1.3.3 Data standardization.....	4
1.3.4 Users apps, browsers & devices.....	4
1.3.5 Users activity .....	4
1.4 Research Questions.....	5
1.5 Research Objectives.....	5
1.6 Research Scope .....	6
1.7 Thesis Organization .....	8
<b>CHAPTER 2 : LITERATURE REVIEW.....</b>	<b>9</b>
2.1 Background.....	9
2.2 Related work.....	10
2.3 Blockchain Technology: Literature review purposes .....	11
2.4 Summary related to this chapter .....	13
<b>CHAPTER 3 : BLOCKCHAIN: AN OVERVIEW.....</b>	<b>15</b>
3.1 Background.....	15
3.1.1 Cryptocurrencies .....	15
3.1.2 Hash function .....	16
3.2 Definition and characterization.....	17
3.3 Key concept.....	18
3.3.1 Transaction (TX) .....	18
3.3.2 Block .....	19

3.3.3 Blockchain .....	19
3.3.4 Mining.....	19
3.3.5 Simple/Normal Node .....	19
3.3.6 Full Nodes .....	20
3.3.7 Miner/Validator Nodes .....	20
3.3.8 TX/Block Finality .....	21
3.3.9 Smart Contracts.....	21
3.3.10 Consensus Protocol.....	22
3.3.11 Consensus Finality .....	22
3.3.12 Proof of work (PoW) .....	23
3.3.13 Proof of Stake (PoS) .....	24
3.4 Types of blockchain.....	<b>25</b>
3.4.1 Public blockchain.....	25
3.4.2 Consortium blockchain .....	25
3.4.3 Private blockchain.....	26
3.5 How does blockchain work? .....	<b>26</b>
3.6 Security analytics .....	<b>27</b>
3.6.1 Tamper proof .....	27
3.6.2 Reliable storage.....	27
<b>CHAPTER 4 : RESEARCH METHODOLOGY .....</b>	<b>28</b>
4.1 Problem formulation .....	<b>28</b>
4.2 Procedure of data collection.....	<b>29</b>
4.3 Proposed architecture.....	<b>29</b>
4.3.1 Data encryption and distribute .....	30
4.3.2 Data decryption and verification.....	31
4.4 Data provider requirements.....	<b>32</b>
4.5 Data requester requirements .....	<b>33</b>
4.6 Summary .....	<b>33</b>
<b>CHAPTER 5 : RESULTS AND DISCUSSION.....</b>	<b>34</b>
5.1 Evaluation .....	<b>34</b>
5.2 Discussion.....	<b>39</b>
<b>CHAPTER 6 : CONCLUSION AND RECOMMENDATIONS .....</b>	<b>41</b>
6.1 Findings and Contributions.....	<b>41</b>
6.2 Recommendations for Future Works .....	<b>42</b>
<b>REFERENCES .....</b>	<b>42</b>

## LIST OF TABLE

Table 3.1: The Comparison of different blockchains .....	25
Table 5.1: Short description of evaluation parameter .....	34
Table 5.2: Comparison between blockchain and traditional database.....	36



## LIST OF FIGURE

Figure 4.1: Preliminary architecture for data sharing platform .....	28
Figure 4.2: Dataset validation and Broadcast process .....	30
Figure 4.3: Data decryption and verification process .....	31
Figure 4.4: Data provider profile management.....	32
Figure 4.5: Data requester goals .....	33

## ABSTRACT

Conventional cloud or central authority based storage has depended solely on extensive capacity suppliers, who act as trusted third parties to exchange and store information. This model represents various issues including information accessibility, high operational expense, and data security. The blockchain is able to maintain transaction without a central authority, the blockchain is an imaginative innovation which opened ways to new applications for tackling various issues in the distributed environment. The primary objectives in this thesis creating secure data sharing platform where the data provider will share their research related data that the verified researcher can use by downloading and contribute to further research. This thesis considers the future prospects of blockchain technology and cryptographic ability in data storage. In this thesis, we proposed an architecture that leverage blockchain technology and provide secure data storage. The data provider shares their data into consortium blockchain network that consist of plaintext and encrypted form the same plaintext hash value via data provider private key (digital signature) and after completion the validating process then the encrypted data will be broadcast into public blockchain network and finally the data requester could request for their desired data via joining the public blockchain network. We have been trying to analysis our proposed architecture performance via comparison among existing blockchain based data storage and traditional database system.

**Keyword: Blockchain, Decentralized storage, Validation.**

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

Blockchain can be barely characterized as a sort of decentralized shared record that utilizes sequential, encrypted and chained squares to store certain and synchronized information over a shared (P2P) network [1]. It creates data with dispersed accord calculations, stores data with encrypted chained blocks, and controls data with self-executed program contents. Blockchain offers a secure, distributed database that can work without a central authority or manager [2]. Blockchain information is put away in each full client hubs taking an interest in a decentralized P2P arrange, what's more, will be confirmed and recorded into blockchain utilizing the accord an instrument [3]. The original of blockchain is an open record for fiscal exchanges with extremely constrained ability to help programmable exchanges [4]. It is an appropriated database that keeps up a persistently developing rundown of exchange data records, cryptographically anchored from altering and correction. The measure of data in our world is quickly expanding. According to an ongoing report, it is assessed that 20% of the world's information has been gathered in the recent years [5]. The decentralization and security qualities of blockchain have pulled in specialists to create different applications, for example, smart contracts, distributed DNS, and character administration and so on [6]. The distributed decentralized data storage will help to reduce the most conventional data failures and blackouts by expanding the security, privacy, and control of the data [7], [8]. A decentralized storage network has been established with many benefit. In

this work we have been trying to pursue a secure place where anyone can share their dataset undoubtedly.

## **1.2 Motivation of the Research**

Digital documents are instrumental in the modernization and rebuild of our current system of putting away and managing printed copies of our essential records. While the advantages are various, computerized documents are additionally open to potential maltreatment and dangers. A huge amount of sensitive data held in data center is helpless against misfortune, leakage, or theft. The current solution incorporates the creation, stockpiling and management of one's records from a single web-based interface. This has drastically enhanced the capacity, recovery, sharing of the reports. Clients have complete control over their information and can have the equivalent with an extensive variety of individuals, including companions, family and different association that requires them for their verification procedure. Numerous such administrations require a lot of capital because of the high expense of building and keeping up a specific data center. This dimension of the foundation must be supported by large third-party organizations.

This existing system is a matter of concern for several reasons. Cloud storage system has always been prone to data theft and loss while also relying too much on a central authority placing full trust on it for storage and maintenance of data [6].

In such a situation of a trustless domain, blockchain rises as the perfect arrangement. Before blockchain, banks had the restraining infrastructure over keeping up exchange records. The development of blockchain democratized this procedure by making an open decentralized record of the equivalent.

Blockchain illuminates the most problem that is begging to be addressed of trust by

wiping out the center man and utilizing a decentralized component to store information transactions. It utilizes cryptographic calculations to avert unapproved access to all records and builds up an entrance control environment where just the individual with the suitable consents can get to the information. Its immutability guarantees that records are tamper-proof and can't be altered once entered further establishing the trust on the system [7], [8].

### **1.3 Problem Statement**

The problem statement in this thesis endeavors to understand is how ready or prepared community currently is about the expected Blockchain revolution. Here are going to be a description about an existing system problem that blockchain has the promising solution to that problem.

#### **1.3.1 Trust on central authority**

The idea of having to trust a central authority for verification is the root cause of the vast majority of the brokenness. Blockchain ignores the requirement for a central authority by distributing information recently decentralized ledger across a network of computers [9].

#### **1.3.2 Potential single point of failure**

A single point of failure (SPOF) is a part of a blockchain system that, on the off chance that it falls, will prevent the whole framework from working. [10] SPOFs are undesirable in any system with an objective of high accessibility or dependability, be it a business practice, programming application, or other modern frameworks.

### **1.3.3 Data standardization**

Data standardization is the critical process of bringing data into a common format that allows for collaborative research, large-scale analytics, and sharing of sophisticated tools and methodologies. Blockchain, originally made as a distributed digital ledger for the original cryptocurrency Bitcoin, Blockchain can be a solution to us petitions, solving data inconsistency problems across vendors, organizations, and customer accomplices [11]. By looking at the core concepts of Blockchain, it's anything but difficult to envision how it very well may be utilized to comprehend information irregularities in showcasing.

### **1.3.4 Users apps, browsers & devices**

They collect information about the apps, browsers, and devices that we use to access their Services or open one of our HTML emails by using different types of technology, including “clear gifs” or “web beacons. “This “automatically collected” information may include Internet Protocol address or other device address or ID, web browser or device type, the web pages or sites that we visit just before or just after the Services, the pages we view on the Services, and the dates and times that we visit the Services.

### **1.3.5 Users activity**

They collect information about our activity in their Services, which may include and views and interactions with content and ads. People with whom we communicate or share content. Which datasets we view or download. It is our presumption that blockchain technology can be a major part in addressing these problems.

## 1.4 Research Questions

**Question 1:** How can we quantitatively characterize data integrity guarantees, in order to enable comparison among different blockchain-based database solutions?

Sensible ways to deal with a reply to this question ought to be founded on the exertion an attacker would spend to compromise data integrity without being recognized.

**Question 2:** How can we design a blockchain-based database with better performances compared to a PoW-based blockchain and with comparable data integrity guarantees?

**Question 3:** What exactly do ensure about the security of proposed architecture?

## 1.5 Research Objectives

- **Objective based on question 1**

**How to Improve Performance?** The performances as of now reachable with PoW-based blockchains are extremely poor as compared with classical database technologies. The tested latency and throughput are relatively contrary with the prerequisites of the thought about cloud situations. In this sense, a testing and crucial research issue respect the examination of novel blockchain plans went for conveying exhibitions adjusted to the present prerequisites while keeping the required integrity guarantees [10]. Quantifying this measure highly depends on the nature of the considered blockchain, but in general a desiderata is that the longer data are stored in a blockchain, the greater the effort an attacker should pay to break data integrity. As a matter of fact, on PoW-based blockchains like Bitcoin's, the attacker effort is close to infinite [11], i.e. it is an infeasible attack.

- **Objective based on question 2**

**How to Enhance Stability?** Current PoW-put together permission less blockchains depend with respect to a market dependent cryptographic money

that may make the putting away of information very costly and excessively needy on market varieties, i.e. it can't guarantee stability. In like manner poor execution cut out 150 numerous conceivable down to earth applications, unsuitable solidness confirmations can seriously confine the pertinence of blockchain to the database. Improving the strength of PoW-based permission less blockchains, e.g. Bitcoin's, adds up to modify the money impetuses basic the mining process. Because of the extensive monetary interests and theory behind digital forms of money, such a change is for all intents and purposes infeasible. A more reasonable way is abusing permissioned blockchains, where incentives don't rely upon digital currencies. The depicted way compares to replying the accompanying research question.

- **Objective based on question 3**

Since, proposed architecture are comprise based on blockchain technology that's why it would be more secure. Because this technology cryptographically secured and for a transaction to be recorded it needs to be verified by nodes. Once recorded it cannot be altered which brings security and stability in the blockchain. In this architecture before accessing plaintext data a data requester have to prove them as a researcher by providing their organizational personal profile url.

## **1.6 Research Scope**

It's rapidly getting to be obvious that blockchain innovation is about much more than just Bitcoin. Recently quick advancements have been occurring in adjusting the Blockchain innovation in different Greenfield regions such as smart energy[12], smart



cities and the sharing economy [13],[14], smart government [15], smart home [16],[17], intelligent transport [18], healthcare[28].

Blockchain technology is a decentralized concept where transactions or smart contracts are maintained in a public ledger which is copied all around the world in different nodes. On a single blockchain, every transaction ever made is available. Anybody can check them using blockchain software's. Transactions are verified by proof of work by miners that check the transactions if they are correct and add them in the next block to be attached in the blockchain.

Blockchain is influencing the market like no one else did. Anybody can use it for any system. With the integration of IOT with blockchain, the system can handle any type of digital goods and communication without any sweat. IoT uses smart contracts and blockchain verifies them with its decentralized network.

Its scope is as wide as internet goes. For any kind of communication, data base, decentralized system will maintain the consistency as well as the verification process, removing the middle men needed for either things.

These days Blockchain techniques are in trending because we are living in a digital world so we need something better for our future.

The blockchain is that future technique where Blockchain help in securing secure-data an ultra-fast as well as it can store data like personal documents, business document or many more things.

## **1.7 Thesis Organization**

The thesis organizations aim to provide direction of every following chapter by grant summary about content. . The first chapter is focusing on the fundamental description of proposed architecture that we will elaborately describe in later chapter. This chapter also hold some keywords such as thesis background, motivation, research scope, and research question. Chapter two the main purpose of findings from the literature review about blockchain technology, accompanying the research approach and related work. The objective of chapter 3 to present an overall concept about blockchain technology. The working procedure in this proposed architecture are narrate in chapter 4. After than chapter 5 present a healthy debate for analysis the proposed architecture result via comparison between blockchain based data storage and traditional data storage. And finally chapter 6 conclude the thesis and recommend for future work.

## CHAPTER 2

### LITERATURE REVIEW

It is followed by a representation of this modified view of the present conditions of research and the literature review provides the technological background of Blockchain and so the base in this paper. It is followed by a representation of the modified version of the conceptual model proposed for this study. The chapter motive to present a theoretical framework and characteristics basis on the topics related to the main research questions and study objective. The research was conducted by using secondary data, such as current incident studies, academic journals and scientific articles version of the conceptual model proposed for this study. Now, we begin by figuring a definition for the essential idea, which is pursued by an introduction of the blockchain technology characteristic. In this chapter, mainly discuss about related area in blockchain technology till now.

#### **2.1 Background**

In its easiest form, blockchain is a publicly distributed ledger or decentralized technology on which valid transactions are anonymously stored. This means the transaction ledger is sustain at one time across a network of unconcerned computers (nodes), and each ledger is replicated thousands of times among the network of nodes [19]. The record contains a ceaseless and finish record (the chain) of all exchanges performed which are gathered into blocks: a block is just added to the chain if the nodes, which are individuals in the blockchain network with elevated amounts of calculation power, achieve consensus on the following 'substantial' block to be added to the chain. Blockchain was imagined by Satoshi Nakamoto in 2008 to serve as public transaction

ledger of the cryptographic money bitcoin [20]. Also, keeping in mind the end goal to decide the legitimacy of a block, "miner" nodes contend to understand an exceptionally complex calculation to verify it (on the Bitcoin Blockchain this is known as the 'proof of Work'). A block for the most part contains four snippets of data: the 'hash' of the past block, a brief of the included transaction, a time stamp, and the Proof of Work that went into making the safe block. When data is entered on the blockchain, it is to a great degree hard to change: It's extremely difficult if any unscrupulous hacker wants to exploit blockchain data because each block includes the previous block hash, so any step to deflect any transaction with the blockchain are simply detectable. Every user has a unique public key made up of an alphanumeric string of 27 to 32 characters that makes it relatively difficult to distinguish the individual it has a place with in this way, while it is not unknown it is pseudonymous [19]. Blockchain can be thought of as an operating systems for which useful applications or "smart contracts" can be written. Resources and data about transactions can be recorded and followed without the contribution of an average middle person, for example, a bank, or a focal expert or some other trusted third party. A blockchain system might be open and public (permission less) like the web or organized inside a private group like an intranet (permissioned) [21].

## **2.2 Related work**

Blockstack [22] presents the idea of virtual chains and proposes a decentralized serverless DNS. Blockstack stretches out to a decentralized public key circulation framework and registry for client identities. Enigma [23, 24] is the nearest to our methodology in that it utilizes the blockchain for access control and empowers sharing of off-chain stored data. However, Enigma stores information get to logs inside the blockchain,

without tending to the noteworthy scalability issues. Additionally, their framework does not suit for IoT stream information. Christian Esposito et al [25] proposed a blockchain based EMR/EHR/PHR ecosystem that allow us to acquire a worldwide view of the patient's medical history in an effective, obvious and lasting way.

For maintaining datasets integrity they propose a reference integrity metric (RIM) for the dataset that is maintained by blockchain. In this methodology there is a central hub that just keeps up references of member repositories where the datasets are really put away and disseminated. There is another chain of block that maintain member information (such as address, sharing policy) and RIM of datasets so that the blockchain guarantee the data integrity [26]. Saqib Ali et al [27] designed a decentralized information storage and access framework for PingER utilizing permissioned blockchain technology. To store health information securely, Huirui Han et al [28] in their proposed model utilize a combined form of consortium and private blockchain named as a hybrid blockchain.

### **2.3 Blockchain Technology: Literature review purposes**

In the easiest form of blockchain technology is a publicly distributed ledger or decentralized technology on which valid transactions are anonymously stored. This means the transaction ledger is sustained at one time across a network of unconcerned computers (nodes), and each ledger is replicated thousands of times among the network of nodes [29]. Blockchain stages cover frequently referred to contestants, for example, Bitcoin, Ethereum, Lite Coin, Block stream, and so on each with their own systems for quick prototyping and application advancement. Bitcoin names itself "an overall cryptographic money and digital payment system" while Ethereum centers intensely on smart contracts self-describing as: "[...] a decentralized stage for applications that run exactly as programmed without any feasibility of fraud, third-party interference”

(Bitcoin, 2016; Ethereum, 2017) [30]. Blockchain technology is a sequentially distributed database where the whole prior exchange history is gathered and shared in a (block) chain in a public ledger. Purposes of this paper, the literature review presented is intentionally summarized, several numbers of systematic literature reviews can be found in various areas [31]. In the recent paper, blockchain technology has plays the role in a different area of the blockchain. Blockchain technology for crypto-currency applications and it was first proposed as a proof-of-work consensus protocol execution of a distributed timestamp server on a decentralized base in the renowned Bitcoin digital currency [32]. Firstly, say about the security of the health information based. The similar informatics system, with their patient information bases, are also nationally or regionally based, with the goal that when the medicinal specialist in one nation or region is required to analyze and treat a guest from some other district or country, she/he should get to the patient's information remotely [32]. If we follow as a security and privacy based field when a healthcare information contains personal and delicate data that may be interesting to cybercriminals. As an example, cybercriminals looking to profit financially from the theft of such information may sell the information to a third-party provider, who may perform data analysis to recognize people who might be uninsurable because of their medical history or genetic disorder [33]. Blockchains are typically shared and synchronized over a peer-to-peer network, and as an example, those are typically used as a public, distributed ledger of transaction records. It is clear that blockchain technology can play a significant role in enhancing the quality of care for patients and possibly decrease costs by more productively allotting resources in terms of personnel, equipment, etc. [33]. Each member in the blockchain system can see the record data also, dismiss or confirm it dependent on an agreement convention [34]. In a recent smart contract are affected by blockchain technology. Now a smart

contracts an automated exchange convention that executes the terms of a contract [35]. The primary targets are to fulfill common contractual conditions, limit special cases both malicious and accidental, and minimize the need for trust. Related economic objectives incorporate bringing down misrepresentation misfortune, discretions and authorization costs, and other exchange costs [33]. The blockchain is a perfect decentralized architecture to ensure distributed transactions between all members in a trustless domain, as IoT systems. In a smart contract, it includes a collection of pre-characterized guidelines and information that have been spared at a particular location of blockchain as a Merkle hash tree, or, in other words, base to-up double tree information structure [36]. Through display in public function or application binary interfaces (ABIs), a smart contract connects with clients to offer predefined business rationale or contract agreement. The smart contract implemented by the security instrument for IoT network has been an interesting issue and some effort have been expressed in a recent example, data protection and access control [37]. An ordinary blockchain, for example, Ethereum depends on the Proof-of-Work protocol, which ensures that token trades happening in the blockchain between individuals are approved by a vast number of nodes that use cryptographic challenges [38]. In this blockchain-based appropriated cloud show, a commitment is some activity that happens outside the blockchain, for example, the execution of a calculation, the exchange of a document, or the arrangement of an arrangement of data, which will lead the event of token exchange between individuals [38].

#### **2.4 Summary related to this chapter**

Based on the paper literature review, Blockchain technology has been defined as its related work and features approach the reader topics at hand and take steps basic knowledge. To be able to discuss the impact of blockchain technology on current

systems, the first structured literature review on the technology, emerged thoroughly on peer-reviewed literature, was observed. Therefore, the technology is desired to have an extensive impact on current and contribute to the construction of new service systems.



## **CHAPTER 3**

### **BLOCKCHAIN: AN OVERVIEW**

The Bitcoin [50] has creatively changed the strategy for financial value exchange with no trusted third party. The basic innovation of Bitcoin is blockchain. In basic terms, blockchain contains a progression of block so that each new block is cryptographically associated with the past block. In the event of Bitcoin, the blocks contain a record of financial TXs between Bitcoin clients. Because of its innate 240 benefits, for example, changelessness, auditability, TX uprightness and verification, adaptation to non-critical failure, or more all without trust task, blockchain is being imagined to assume an indispensable job in the security of IoT biological community.

#### **3.1 Background**

##### **3.1.1 Cryptocurrencies**

Coinmarketcap.com is a site that lists distinctive cryptographic forms of money. So far coin showcase top records more than 1300 diverse digital forms of money. Typically, every money has their very own motivation what's more, use. The primary cryptographic forms of money were centered around peer-to-peer payments, in any case, later on, there has been an ascent in monetary forms which accomplishes something beyond that. There are monetary forms for following of drug store items with IoT devices like Modum [39], and there are currencies like Dentacoin [40] which centers on enhancing dental consideration around the world, making it moderate through group control, the cash Tron

means to construct a tipping administration for online networking content, stimulation substance, and spilling [41]. Monetary standards like Bitcoin, Litecoin, Bitcoin Cash, Dash and Monero and others, center around decentralized peer-to-peer payments with each having their own one of a kind usage of the blockchain. Monetary forms like Ethereum, Cardano, NEO, RootStock, and Namecoin to take into consideration decentralized peer to peer, yet in addition centers on the use of smart contracts, which enables their clients to store code on their blockchain and have it kept running for a charge.

Numerous digital forms of money utilize the blockchain here and there or another; be that as it may, there are special cases. IOTA, for example, utilizes a direct acyclic graph (DAG) with the end goal to do exchanges [42]. The monetary standards typically have an encompassing framework empowering the usefulness they guarantee to serve. A portion of the monetary standards is coins, implying that they have their exceptionally claim blockchain, a portion of the monetary forms are tokens which rely upon another monetary form stage to work, as a rule being Ethereum. These monetary forms' tokens or coins are exchanged at an open market and costs for each can be seen at [coinmarketcap.com](https://coinmarketcap.com) [43].

### **3.1.2 Hash function**

For the most part, a hash work takes any information of any length as information and change the information with a deterministic mathematical function. A hash capacity can be utilized to outline of discretionary size to information of fixed size. The yield of a hashing capacity can be mapped back to the first contribution, without essentially giving us a chance to reproduce the first item. Unique hashing calculations will give distinctive yield with a wide

range of properties. For precedent in seeking, hashing may accelerate such process by ordering information in a cluster, at that point scan for the hashed string in the cluster with hashed information will show where the information is put away in the exhibit [44] [45]. A cryptographic hash capacity can give affirmation of information trustworthiness. A cryptographic hash capacity can be utilized to make a unique finger impression of a few information. This is helpful in the manner in which that you can check the unique mark of the information, to guarantee that the information has not been altered. On the off chance that the information has been messed with, the unique mark won't be substantial. This additionally applies when the information isn't put away safely, you will dependably have the capacity to recalculate the fingerprint of the information, to check on the off chance that it coordinates the old finger impression. In the event that the fingerprints coordinate, the respectability of the information isn't disabled in any capacity. A cryptographic hash work isn't equivalent to encryption. A cryptographic hash work just makes a "process" of the first information in a restricted capacity, making recovery of the first information unrealistic. In asymmetric encryption work, one will have the capacity to recover the original data [46].

### **3.2 Definition and characterization**

Blockchain is a rising technology that empowers new structures of disseminated software designs, where ingredients can find consent on their shared states for decentralized and transactional information sharing over a huge network of untrusted members without depending on a focal integration point that should be reliable by every elements inside the framework [47]. Blockchain and distributed ledger innovation offers huge and adaptable handling power, high exactness rates, and clearly

unbreakable security at a fundamentally diminished cost contrasted with the conventional frameworks the innovation could supplant, for example, settlement, trading and accounting system [48].

A blockchain basically chain of block, is increasing list of records, called block, which are connected using cryptography. Each block comprises a cryptographic hash of the former block, a timestamp and transactional data [49]. By plan, a blockchain is impervious to alteration of the information. It is "an open, disseminated record that can record exchanges between two parties productively and in a verifiable and changeless way. The creation of the blockchain for bitcoin made it the principal advanced for digital currency to take care of the double spending issue without the need of a trusted authority or central server. A blockchain is a decentralized, disseminated and open computerized record that is utilized to record transactions crosswise over numerous PCs so the record can't be modified retroactively without the modification of every single consequent block and the consensus of the system [49].

### **3.3 Key concept**

#### **3.3.1 Transaction (TX)**

A procedure that outcomes in the difference in the condition of the blockchain. Contingent on the blockchain stage, a TX ranges from the exchange of a financial incentive to the execution of a self-assertive code as a smart contract [50].

### **3.3.2 Block**

It is a set of TXs that happened in the recent past and have not been confirmed yet. The block also has a block header that contains, blockchain version number, hash of the previous block, a random nonce, time stamp and Merkle Root Hash of all the TXs included in the block.

### **3.3.3 Blockchain**

It is a distributed public ledger that keeps a record of all the TXs/blocks. Vitalik Buterin in [51] gives another perspective that the essence of the blockchain is informational and processual, and does not relate directly to the monetary sphere.

### **3.3.4 Mining**

It is the procedure of adding validated TXs to a block and then broadcasting that block on the blockchain network, to be known by all the nodes. The mining is done by miner nodes, and the selection of a node to mine a new block is done based on certain lottery schemes. In case of Bitcoin, miners compete to solve a cryptographic hash puzzle and whosoever finds the solution (also known as proof of work) first, is eligible to mine the next block. When a block is mined and added to the blockchain, then the TXs in that block are confirmed [52]. Irrespective of the type of blockchain platform, usually some lottery scheme is required to randomly select a miner to propose or mine a new block.

### **3.3.5 Simple/Normal Node**

There may be several types of nodes in a blockchain network reline upon their capabilities and resources such as computation capability and memory size. A node may be a simple node, which can only send and receive a TX and does not

store the complete copy of the blockchain. In case of an IoT environment a simple node can be an Arduino-based sensor node that can only send a sensor reading to the gateway device or receive some commands.

### **3.3.6 Full Nodes**

These nodes keep up an entire duplicate of the blockchain, however, they don't mine a block. However, full nodes approve TXs dependent on the agreed standards of the individual blockchain and contribute in tolerating or forking out a square [53]. A twofold spending or a malignant TX may not be steered or transferred by a full-node. This suggests full nodes are fit for TX and block engendering. Consequently, full nodes are fundamental for the security of the blockchain. In an IoT domain a Raspberry Pi (Rpi) with increasingly computational and memory assets when contrasted with an Arduino, can be a full hub [54]. It was likewise tried by running a Go Ethereum form geth-Linux-arm7-1.8.3 on a Rpi-3 based sensor node.

### **3.3.7 Miner/Validator Nodes**

These are the full nodes that have the additional capability to mine or validate a new block thus extending the blockchain [53]. Moreover, mining nodes are selected as per specific criteria based upon the type of consensus protocol being used in the blockchain. E.g., In Bitcoin, the mining nodes have to solve a cryptographic puzzle, and the node that does it first is eligible to mine the block. The miner node has to submit a Proof of Work (PoW) along with the mined block so that the rest of the nodes can validate that the puzzle has been correctly solved. If the block is accepted by the rest of the network, the miner node then earns a block 280 reward and TX fee in the form of respective cryptocurrency.

Whereas, in Proof of Stake (PoS) consensus protocol, miner nodes are selected randomly based on the coinage, i.e., the number of coins they own and the time since they have those coins. However, in most of the (Byzantine Fault Tolerance) BFT-based consensus protocols, the validator is elected in a round robin fashion to propose a new block. The rest of the member nodes of the quorum, vote on the validity of the block and its TXs. In most of the cases, the block is validated and included in the blockchain upon getting 2/3 majority votes in its favor.

### **3.3.8 TX/Block Finality**

It is related to the final confirmation or approval of a particular TX or a block by the consensus protocol of respective blockchain. It is an important aspect as it infers delay in TX confirmation and ultimately affects the TX throughput of the blockchain. E.g., In Bitcoin, a TX gets one confirmation/approval after 10 minutes, i.e., once the block containing that TX is mined. However, to get a final confirmation, the TX has to wait until additional five blocks are mined and appended to the block containing that particular TX. Hence, it takes 60 minutes to finally declare a TX confirmed/approved in Bitcoin blockchain. Whereas, in other blockchains such as Hyperledger [54] and Tendermint [55] the TX gets instant confirmation.

### **3.3.9 Smart Contracts**

Exploiting the Bitcoin's ability to execute autonomous scripts, developers have created new versions of the blockchain that can perform arbitrary computations other than transferring coins. E.g., Ethereum blockchain implements scripts called smart contracts that can run any algorithm encoded in them as a part of

the TX [56]. Being deployed on the blockchain, the smart contracts are also called as Decentralized Applications or DApps". Since smart contracts reside on the blockchain, they have a unique address. A smart contract can be triggered by addressing a TX to it under some rules that govern the contract. Smart contracts can be used in applications like auto-pay (shopping, parking, route management, tolls, fuel payment), digital rights management, financial services including loan, inheritances, escrow, cryptocurrency wallet controls, capital markets, mortgage, automatic payment of insurance claims [57], SCM and smart grid [57]. The key idea behind smart contracts is the development of autonomous objects or IoT devices that cannot only rent or sell their data but also maintain their operability by paying for the maintenance services. Such an autonomous system is likely to contribute to the development of an overall "Economy of Things" with the goal of providing efficient and consistent services without any intermediary.

### **3.3.10 Consensus Protocol**

It is the mechanism or set of rules that enables all the full nodes to reach an agreement over the order of TXs. There are many types of consensus protocols being used in different blockchain applications. E.g., PoW, PoS, Practical Byzantine Fault Tolerance (PBFT), etc. Some of the notable consensus protocols are being discussed in succeeding paras.

### **3.3.11 Consensus Finality**

It means, the convergence of the blockchain consensus process on a particular block/order of TXs. However, in reality, a consensus process may result into a permanent block or a stale block that may be forked out later. This aspect is



further illustrated by Vitalik Buterin in [53], that the finality of a TX is always probabilistic. However, it may stand true for a PoW, PoS or PoET consensus protocol [58], but other consensus protocols may have different finality guarantees. Such as Casper [53] offers stronger finality guarantees as compared to PoW consensus and similarly, BFT-based consensus protocols provide immediate consensus finality [58, 59], and the TXs once confirmed are not forked out later. From IoT point of view consensus finality is an essential requirement in most of the IoT systems as it also influences latency in TX confirmation.

### **3.3.12 Proof of work (PoW)**

It is the calculation of a cryptographic hash work with some level of difficulty [48], i.e., choosing a nonce to such an extent that the processed cryptographic hash has an explicit number of zeros in the begin as characterized by the dimension of trouble. PoW frames the premise of agreement strategies in Bitcoin and different digital forms of money. At the point when a mineworker hub unravels the PoW, it is qualified to mine another square. Though, other full nodes in the system commonly affirm its accuracy [50]. PoW secures against twofold spending assaults. Since it is computationally serious, it is trying for a solitary aggressor to illuminate the trouble for all the adjusted squares previously the legit nodes in the system [42]. It is a typical discernment that if a noxious excavator or a pool of mineworkers increase 51% of the aggregate system hash control, they can control the system [65]. Be that as it may, creators in [63] demonstrate that the vindictive/untrustworthy mineworkers turning to narrow-minded mining technique can acquire income by just 25% of the aggregate hashing power. Along these lines, least  $2/3$  of the system nodes

should be straightforward to ensure against narrowly minded mining; a basic lion's share isn't sufficient. Also, open systems with pseudonymous client IDs are inclined to Sybil assault. In this manner, Satoshi Nakamoto imagined PoW-based accord for Bitcoin blockchain to make Sybil assaults progressively costly to be propelled [63, 65].

### **3.3.13 Proof of Stake (PoS)**

It was conceived based on an idea described in [65] to improve upon PoW's high latency, high computation, and high energy costs. PoS implies that the people with high stakes are less likely to attack the respective network. Hence, an entity with the highest coinage, i.e., number of coins times the days, will be eligible to mine a new block. Moreover, the mining difficulty is inversely proportional to the coinage [8]. However, once the miners claim the reward, the coinage is reset so that other miners/stakeholders also get the chance to mine a block. Therefore, if an attacker wants to launch an attack similar to 51% attack, he must own enough coins so that even when the coinage is reset, he can still gain more than half of the odds [8]. In addition, Nicolas Houy in [65] proves that PoS is vulnerable to a 51% attack, as the few rich stakeholders can collude to manipulate the state of the ledger. Nevertheless, the probability of a 51% attack in PoS is considered to be lower as compared to the PoW [65]. Moreover, the maximum TX rate a PoS protocol has achieved is a few hundred TPS (Transactions per Second) as compared to Visa's peak capacity of 56000 TPS [66, 67]. Due to the lack of consensus finality, PoS-based consensus can also lead to blockchain forks [66]. A variation of PoS named "Delegated Proof of Stake" (DPoS) [68, 69] implemented in Bitshare, a digital currency, is

considered to be more efficient than PoS in terms of TX confirmation time. Moreover, it can tolerate up to 50% malicious nodes [43, 69].

### **3.4 Types of blockchain**

The name of blockchain comes from its technical structure a chain of the block. In this paper, we are going to show the different types of Blockchain: Permission less Blockchains, Permissioned Blockchains. Every block is connected to the previous block with a cryptographic hash. A block is a data structure which agrees to store a list of transactions [70]. Since understanding some of the basic rules of the technology, it is time to more characterize between different types of Blockchain [71]. Blockchains are at that point ordinarily classified in two primary classifications: We are going to indicate to the different types of Blockchain: Public blockchain, consortium blockchain and fully private blockchain [73], [74]. The comparison of different blockchain is listed in Table 3.1.

#### **3.4.1 Public blockchain**

The based on public Blockchain protocols are open source by Proof of Work (PoW) consensus algorithms and anyone can participate, without permission.

#### **3.4.2 Consortium blockchain**

Consortium blockchain shows a blockchain where consensus process is controlled by pre-chosen set of nodes. For instance, there is a consortium of 15 monetary foundations, every one of which works a node and of which 10 organizations need to approve the new block to influence it to be legitimate.

### 3.4.3 Private blockchain

In a Private Blockchain, alter permissions are kept centralized to one organization. Private blockchain are a method for exploiting blockchain technology by setting up groups and members who can check transactions internally.

**Table 3.1.** The Comparison of different blockchains

Property	Public blockchain	Consortium blockchain	Fully private Blockchain
Consensus determination	All nodes	Selected nodes	A specific institution
Read Permission	Public	Could be public or restricted	Could be public or Restricted
Efficiency	Low	High	High
Centralized	No	Partial	Yes

### 3.5 How does blockchain work?

A blockchain is the structure of data that illustrate a financial record entry, or a record of a transaction. Every transaction is digitally signed to assure its authenticity. When a current transaction or an edit to an existing transaction comes into a blockchain, usually a majority of the nodes into a blockchain implementation must execute algorithms to appreciate and verify the history of the individual blockchain block that is proposed [75]. If a majority of the nodes appear to a unity that the history and signature are valid, the new block of transactions is accepted into the ledger and a new block is added to the chain of transactions. If a majority does not confess to the connection or modification of the ledger entry, it is identified and not added to the chain.

### **3.6 Security analytics**

This section presents a security analytics of public blockchain that we used to share dataset. By comparing public blockchain and other strategies in recent researches, we will try to analyze the advantages of public blockchain.

#### **3.6.1 Tamper proof**

Basically, tamper proof of data is effective in the blockchain by two things; a cryptographic fingerprint that are unique each block and a consensus protocol which process a deal to add new block in the network. The fingerprint called hash that require huge computing power and time to generate initially and linked each block with this hash. So, a hacker compete against the world wide computing power that makes very difficult to tamper the data. But it would be possible if 51% miner agreed to rewrite.

#### **3.6.2 Reliable storage**

Decentralized capacity is relied upon to bring the best properties of Blockchain innovation together. The features of decentralized stockpiling will take care of the down to earth demand of putting away huge measures of information. Any participants are able to verify data using cryptographic hash value at any case.

In our proposed architecture the data provider has the ownership of his/her data and the control of using the data. Since, all data are stored in a distributed way the same dataset are kept in every node. It solves the issue of single point of failure.

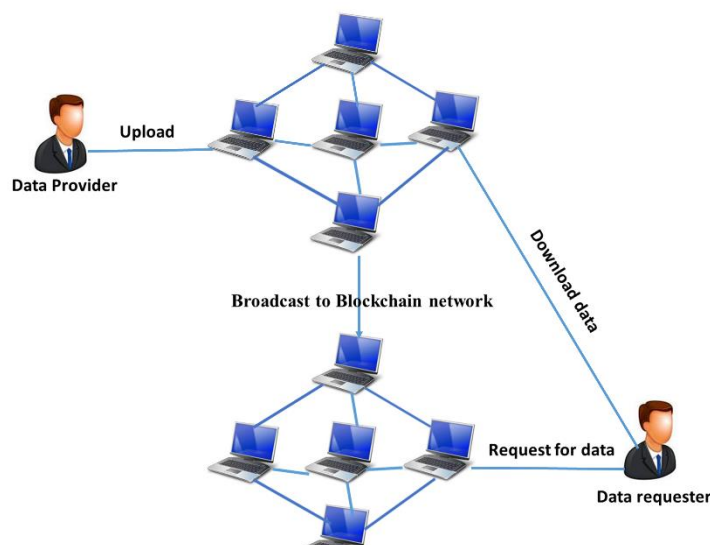
## CHAPTER 4

### RESEARCH METHODOLOGY

#### 4.1 Problem formulation

Our proposed architecture has three general part: Data provider, Data requester and Blockchain network as demonstrated in Figure 4.1.

- **Data provider:** Data provider wants to share/store their data in a secure place where they could save the cost of the data storage and maintenance. Here, they will create their data profile.
- **Data requester:** A data requester is a subscriber for the dataset that uploaded by data owner. Data requester have to join with blockchain network as a node and then prove themselves as a researcher by providing organizational mail, organizational personal profile url for the purpose of downloading required dataset.
- **Blockchain network:** This network is the core concept for this proposed architecture.



**Figure 4.1:** Preliminary architecture for data sharing platform

- **Workflow:** In our proposed architecture we use consortium blockchain network for validating the data and then just encrypted hash value will be broadcast to the public blockchain network via communication network. For better understand we consider an example; A data provider Alice want to share her data to the blockchain network. In order to ensure data integrity and data security she needs to follow some procedure before uploading dataset, in the later section we will describe those procedure. Bob as a data requester firstly he will be join with the public blockchain network and then search for dataset with required keyword and this keyword match with the data profile keyword. If he find the desired dataset then he create a data request but before access the dataset validation node verify requester identity as a researcher based on predefined rules (smart contract) that execute automatically.

#### **4.2 Procedure of data collection**

Mainly this proposed architecture provide a secure decentralized storage facilities where anyone can share their dataset. For sharing their dataset they have to follow some procedure. A user can share their dataset in the consortium blockchain network by themselves. After validate those data, broadcast to the public blockchain network. A data provider would have to share hash form of data with digital signature and plaintext of the same data for the purpose of data verification.

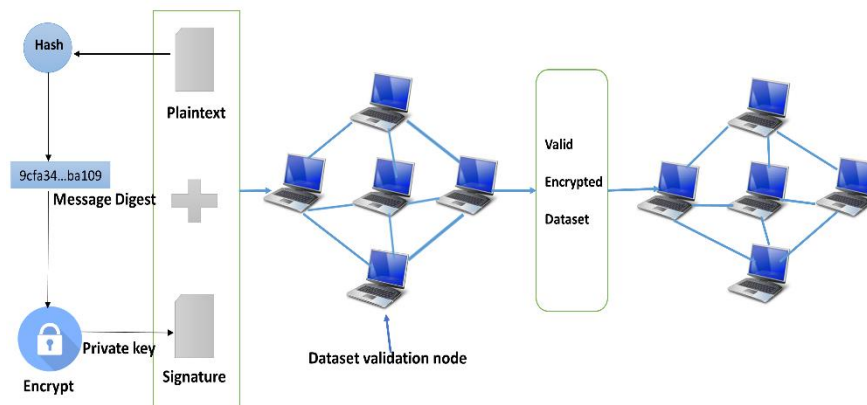
#### **4.3 Proposed architecture**

In this proposed architecture, all gathered data are eventually recorded in a digital ledger in the form of linked blocks which remain in distributed form. It is very important

to maintain data accuracy, data broadcast, and data verification, via data validation process. In this section we illustrate in detail of the working procedure of the proposed architecture which build mainly of data transmission, verification and record.

### 4.3.1 Data encryption and distribute

Each validation-node in the consortium blockchain network have a pair of public and private key. The public key that is available publicly in the validation-node network. The private key that is used to verify node's identity and operation that it may perform. Since it is a blockchain based distributed network therefore each node get encrypted data and share to other node in the network. The data encryption and distributed process are illustrated in Figure 4.2.



**Figure 4.2:** Dataset validation and Broadcast process

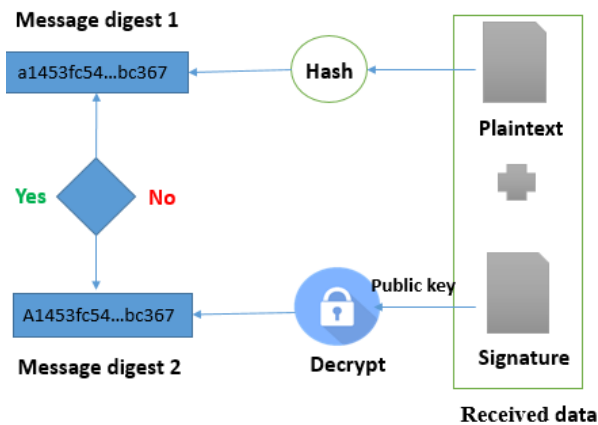
The sharing data are consist of plaintext and digital signatures. In the encryption mechanism the new plaintext data are refined using secure hash algorithm (SHA), produce a message digest. The private key of each validation-node encrypt the message digest forming digital signature which decrypt using its



public key. Then the sharing data is broadcast to all other validation-node via the communication network.

### 4.3.2 Data decryption and verification

All validation-node that are received broadcast information need to decrypt the receive information and verify the result as depicted in Figure 4.3.

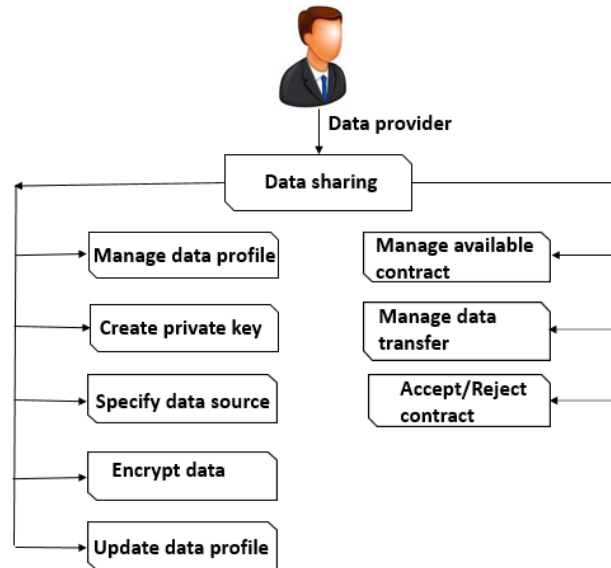


**Figure 4.3:** Data decryption and verification process

Again the plaintext data should proceed using same hash algorithm and generating message digest 1 on the contrary decrypt the signature using its public key and generate message digest 2. Now compare both digest 1 and digest 2, if both digest are same the received information is successfully verified; otherwise received information considered as false. After completion both process valid data are stored to blockchain based distributed ledger.

#### 4.4 Data provider requirements

In Figure 4.4 Represent the data provider core elements-governing the data profile and obtainable contracts. For the purpose of storing data securely in blockchain network a data provider needs to maintain a data profile that include private key management and

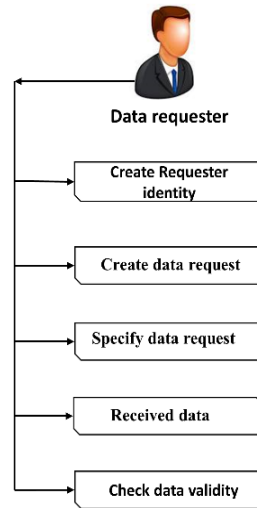


**Figure 4.4:** Data provider profile management

storage which is used to encrypt the message digest thereby forming a digital signature. This digital signature verify the data provider identity. The data profile needs to link with the data sources via API access keys that the data profile should be updated. The data provider also require to provide personal identity source such as organizational profile url, organizational mail.

## 4.5 Data requester requirements

The data requester core requirements are illustrated in Figure 4.5.



**Figure 4.5:** Data requester goals

The requester have to join the public blockchain network for searching required dataset after that create data request for specific data. This request notification will notify the consortium blockchain network before release the dataset the validator verify the requester identity as a researcher using requester organizational mail, organizational profile url. If the requester successfully verified by the validation node and then the requester can download desired dataset.

## 4.6 Summary

This chapter introduced the concept of our proposed architecture which is the major service offered by the blockchain technology. This chapter stated that how a data provider share their data into blockchain network and how a data requester access their required data. In the next chapter we will try to analysis our proposed architecture performance by uphold a comparison between blockchain and legacy database.

## CHAPTER 5

### RESULTS AND DISCUSSION

The results of this investigation revealed that the use of Blockchain for Data storage is mostly in its conception stage. The quantity of amazing papers identified with this combined innovation is small; nonetheless, it is extremely noticeable that the data scientist is very keen on this topic. The current research of protecting data on Blockchain is concentrated on finding and distinguishing enhancements to the current difficulties and confinements. White papers, specialized discussion papers from topic specialists on the related subject and industry reports are the primary source of information on this issue [76], [77], [78].

#### 5.1 Evaluation

Here, we have been given a short description of evaluation parameter and later that we use this parameter to comparison between blockchain and traditional database. Table 5.1 shows the definition of evaluation parameter.

**Table 5.1:** Short description of evaluation parameter

Evaluation parameter	Definition
Transparency[61]	Transparency, as utilized in science, engineering, business, the humanities and in other social settings, is working so that it is simple for others to perceive what activities are performed.
Integrity[61][85]	Integrity is the nature of being straightforward and having a solid good principle or good uprightness or accuracy of one's action.

Performance[86]	Here, performance indicates the time taken by blockchain or database to finish a task efficiently or finishing an assignment without expending much time.
Data ownership[87]	Data ownership is the demonstration of having lawful rights and full control over a solitary piece or set of information components.
Access control[88]	Access control is a security method that manages who or what can view or utilize assets in a computing domain. It is a key idea in security that limits risk to the business or organization.
Trust[66]	Meanings of trust regularly refer to a circumstance described by the following viewpoints: One party will depend on the activities of another party.
Database validity[89]	Database validity guarantees that all wrong information is excluded from the database.
Privacy and security[90]	Privacy guarantees that individual data (and some of the time corporate private data too) are gathered, prepared (utilized), protected and obliterated lawfully and decently. ... Security controls limit access to individual data and ensure against its unapproved utilize and obtaining.
Concurrency and synchronization[54]	In software engineering, concurrency indicates the capacity of various parts or units of a program, calculation, or issue to be executed out-of-order or in halfway request, without influencing the ultimate result. Information synchronization advancements are intended to synchronize a single set of

	information between at least two devices, consequently replicating changes forward and backward.
Reliability and availability[56]	Reliability speaks to the likelihood of components, parts, and frameworks to play out their required functions for the desired timeframe without failure in determined situations with the desired certainty. Availability is characterized as the likelihood that the system is working appropriately when it is asked for utilizing.
Transaction creation	Transaction creation means how data transfer one user to another user and this transfer data recorded as a transaction.
Fraudulent/malicious change[91]	Fraudulent/malicious change indicates which steal protected data, delete documents or add software not approved by a user.

To evaluate the proposed blockchain architecture for dataset sharing platform, we compared it with traditional database system [79], [80], [81], [82] [83], [84]. Table 5.2 shows the comparison.

**Table 5.2:** Comparison between blockchain and traditional database

Evaluation parameter	Blockchain	Traditional Database
Transparency	All the user can see how the blockchain work with time	On the contrary traditional database provides information only after user credentials are authenticated

Integrity	The user can be sure that the data remain unaltered and uncorrupted from the time it was recorded	Traditional database cannot give the guarantee that data remain unaltered
Performance	Blockchain considered as slow database though it is ideal as transaction platform.	Since the traditional database has been existence for decades now. There has been increase in performance.
Data ownership	Maintained by cryptographic key pairs and native cryptographic algorithm	Established through central authority
Access control	Inherently identical for all permissioned nodes	Centrally administrated
Trust	Native via immutable records	Established via central authority
Database validity	Continuous	Provided only for single instances In time

Privacy and security	Cryptographic authentication	Each row based enactment from central authority
Concurrency and synchronization	Consensus yields identical copy.	Through complex checking between central DB and user DB to ensure agreement
Reliability and availability	Peer to peer network to distribute data replication across all nodes	Potential single point of failure
Transaction creation	Available for all permissioned nodes	Managed via central authority
Fraudulent/malicious change	Almost difficult because all block are connected through cryptographic hash	Not available where current keys and check constraint remain Insufficient



## 5.2 Discussion

The key territory of blockchain is 'open check' which is empowered because of integrity and transparency work. This helps the clients or customers to check every one of the points of interest on the block. They are completely mindful about what is happening in the blockchain. The traditional database despite provides details to clients or customers simply after their accreditations are verified. The authority gives just the required data and not all that matters [5]. From this comparison undoubtedly we can say that user depend on blockchain based proposed architecture than traditional database. Because a user can see what happened within the network. In this proposed architecture maintain data integrity using cryptographic hash function. Once the data stored into the network nobody can altered these data because the data would be cryptographically secured using SHA256 hash algorithm [3] further more data validation process helps to maintain data integrity (Chapter IV). The traditional database on the contrary does not guarantee the data integrity because the administrator can alter the data without anyone knowing. For performance analysis blockchain remain its infancy stage [5], [6]. Blockchains are considered slow databases, even though they are systems of records and are ideal as transaction platform. We know that the blockchain is improving day after day, but the nature of the technology requires sacrifice in terms of speed. It's totally different with regards to the traditional database. The database has been advancing over the timeframe and has been in existence throughout recent decades. There has been increase in performance [7]. A comparison among Blockchain and traditional Databases (See Table 2) uncovers a great deal of fascinating information. The most essential point of distinction is the lacking of central authority which provides to the other differences. Ownership for information is set up by means of cryptographic key pairs in blockchain and by the central authority on

account of traditional databases. For blockchain's situation, data legitimacy is guaranteed for the whole chain while the traditional databases just give it to single cases in time. The consensus characteristics of blockchains keep up indistinguishable duplicates with all clients while traditional databases require complex checking between the central database and client database for ensuring agreement. All permissioned clients inherently have indistinguishable access with blockchain. Peer to peer networking for decentralized distributed data copy doesn't allow for a single point of failure. Since each blocks are dependent on the previous blocks, blockchains are secured against false changes while legacy databases suffer due to the inefficiency of keys and constraints [78], [80], [83], [84]. From this discussion we can easily understand that blockchain based proposed architecture would be reliable platform for researchers to share their research related data without concern about data forgery.

## CHAPTER 6

### CONCLUSION AND RECOMMENDATIONS

#### 6.1 Findings and Contributions

Here we will try to uphold some existing research findings that we have been trying to mitigate those findings through our proposed architecture. In this article [3] they proposed a private blockchain based data protection framework that store encrypted data alongside plaintext data. I think anyone can access those plaintext data by joining their network. But in our proposed architecture we used two types of blockchain network as consortium we used just verified the data and then only encrypted hash value are stored into public blockchain network. Our consortium blockchain network will consist of predefined nodes that validate the data. Peer to peer connection does not support single point of failure if any node goes down then the system continue work where in traditional database if any single point failure occurred then the system will stop working.

Xia et al. [4] in their research they proposed BBDS system that used permissioned blockchain which allows access to only invited and verified users. That means if a user wants to access their data then the user should have to prove their identities and cryptographic key pairs. There is the main problem, after successfully completion user authentication if the user don't find their required data then the whole authentication process that already happened are useless. To mitigate this problem in our proposed architecture we use public blockchain where anyone can search their required data in encrypted form, and then they could create a data request. After that verify the requester identity as a researcher, if the requester verified successfully as a researcher then the data

could be released. Verified data requester could download data from consortium blockchain network.

## 6.2 Recommendations for Future Works

In this proposed decentralized distributed data sharing platform architecture uses Consortium and permission less (public) blockchain. This proposed architecture removes the centralized repository. Using blockchain technology this architecture also ensure the data integrity via cryptographic key pair and hash function. By actualizing the proposed model, clients will have the capacity to store all their documents on an online medium with no danger of information leakage. It is fascinating to expand this work by completely investigating these in future studies. In our future work should address implementation on our proposed architecture.

## REFERENCES

- [1] “*Intro to Blockchain Technology*”,<https://blockchainhub.net/blockchain-technology>, last accessed: 18 October 2018
- [2] Angraal, S., Krumholz, H. M., & Schulz, W. L. (2017). Blockchain technology: applications in health care. *Circulation: Cardiovascular Quality and Outcomes*, *10*(9), e003800.
- [3] Yuan, Y., & Wang, F. Y. (2016, November). Towards blockchain-based intelligent transportation systems. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on* (pp. 2663-2668). *IEEE*.
- [4] Xu, X., Pautasso, C., Zhu, L., Gramoli, V., Ponomarev, A., Tran, A. B., & Chen, S. (2016, April). The blockchain as a software connector. In *2016 13th Working IEEE/IFIP Conference on Software Architecture (WICSA)* (pp. 182-191). *IEEE*.
- [5] Zyskind, G., & Nathan, O. (2015, May). Decentralizing privacy: Using blockchain to protect personal data. In *Security and Privacy Workshops (SPW), 2015 IEEE* (pp. 180-184). *IEEE*.

- [6] Liang, X., Shetty, S., Tosh, D., Kamhoua, C., Kwiat, K., & Njilla, L. (2017, May). Prochain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability. *In Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (pp. 468-477)*. IEEE Press.
- [7] Kuo, T. T., Kim, H. E., & Ohno-Machado, L. (2017). Blockchain distributed ledger technologies for biomedical and health care applications. *Journal of the American Medical Informatics Association, 24(6), 1211-1220*.
- [8] Shafagh, H., Burkhalter, L., Hithnawi, A., & Duquennoy, S. (2017, November). Towards blockchain-based auditable storage and sharing of iot data. *In Proceedings of the 2017 on Cloud Computing Security Workshop (pp. 45-50)*. ACM.
- [9] Kuo, T. T., & Ohno-Machado, L. (2018). ModelChain: Decentralized Privacy-Preserving Healthcare Predictive Modeling Framework on Private Blockchain Networks. *arXiv preprint arXiv:1802.01746*.
- [10] Gaetani, E., Aniello, L., Baldoni, R., Lombardi, F., Margheri, A., & Sassone, V. (2017). Blockchain-based database to ensure data integrity in cloud computing environments.
- [11] Garay, J., Kiayias, A., & Leonardos, N. (2015, April). The bitcoin backbone protocol: Analysis and applications. In Annual International Conference on the Theory and Applications of Cryptographic Techniques (pp. 281-310). *Springer, Berlin, Heidelberg*.
- [12] Li, J., Greenwood, D., & Kassem, M. (2018, June). Blockchain in the built environment: analysing current applications and developing an emergent framework. *Diamond Congress Ltd*.
- [13] Rivera, R., Robledo, J. G., Larios, V. M., & Avalos, J. M. (2017, September). How digital identity on blockchain can contribute in a smart city environment. *In Smart Cities Conference (ISC2), 2017 International (pp. 1-4)*. IEEE.
- [14] Biswas, K., & Muthukkumarasamy, V. (2016, December). Securing smart cities using blockchain technology. In High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2016 IEEE 18th International Conference on (pp. 1392-1393). IEEE.
- [15] Hou, H. (2017, July). The application of blockchain technology in E-government in China. In *Computer Communication and Networks (ICCCN), 2017 26th International Conference on (pp. 1-4)*. IEEE.
- [16] Dorri, A., Kanhere, S. S., Jurdak, R., & Gauravaram, P. (2017, March). Blockchain for IoT security and privacy: The case study of a smart home. In

Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on (pp. 618-623). IEEE.

- [17] Lazaroiu, C., & Roscia, M. (2017, November). Smart district through IoT and Blockchain. In Renewable Energy Research and Applications (ICRERA), 2017 IEEE 6th International Conference on (pp. 454-461). IEEE.
- [18] Yuan, Y., & Wang, F. Y. (2016, November). Towards blockchain-based intelligent transportation systems. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on* (pp. 2663-2668). IEEE.
- [19] Li, J., Greenwood, D., & Kassem, M. (2018, June). Blockchain in the built environment: analysing current applications and developing an emergent framework. *Diamond Congress Ltd.*
- [20] Blockchain, <https://www.forbes.com/sites/bernardmarr/2018/02/16/a-very-brief-history-of-blockchain-technology-everyone-should-read/#8c15937bc470> Accessed on 11-09-2018.
- [21] BlockchainBackground, <https://www.dlapiper.com/en/denmark/insights/publications/2017/06/blockchain-background-challenges-legal-issues/> 11-09-2018
- [22] Ali, M., Nelson, J. C., Shea, R., & Freedman, M. J. (2016, June). Blockstack: A Global Naming and Storage System Secured by *Blockchains*. In *USENIX Annual Technical Conference* (pp. 181-194).
- [23] Zyskind, G., & Nathan, O. (2015, May). Decentralizing privacy: Using blockchain to protect personal data. In Security and Privacy Workshops (SPW), 2015 IEEE (pp. 180-184). IEEE.
- [24] Zyskind, G., Nathan, O., & Pentland, A. (2015). Enigma: Decentralized computation platform with guaranteed privacy. *arXiv preprint arXiv:1506.03471*.
- [25] Esposito, C., De Santis, A., Tortora, G., Chang, H., & Choo, K. K. R. (2018). Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?. *IEEE Cloud Computing*, 5(1), 31-37.
- [26] Banerjee, M., Lee, J., & Choo, K. K. R. (2018). A blockchain future for internet of things security: a position paper. *Digital Communications and Networks*, 4(3), 149-160.
- [27] Ali, S., Wang, G., White, B., & Cottrell, R. L. (2018, August). A Blockchain-Based Decentralized Data Storage and Access Framework for PingER. In 2018 17th

- [28] Han, H., Huang, M., Zhang, Y., & Bhatti, U. A. (2018, June). An Architecture of Secure Health Information Storage System Based on Blockchain Technology. In International Conference on Cloud Computing and Security (pp. 578-588). Springer, Cham.
- [29] Li, J., Greenwood, D., & Kassem, M. (2018, June). Blockchain in the built environment: analysing current applications and developing an emergent framework. *Diamond Congress Ltd.*
- [30] PROWSE, S. (2017). BEYOND BITCOIN: A LITERATURE REVIEW OF BLOCKCHAIN TECHNOLOGY.
- [31] Lindman, J., Tuunainen, V. K., & Rossi, M. (2017). *Opportunities and risks of Blockchain Technologies—a research agenda.*
- [32] Kuo, T. T., & Ohno-Machado, L. (2018). ModelChain: Decentralized Privacy-Preserving Healthcare Predictive Modeling Framework on Private Blockchain Networks. *arXiv preprint arXiv:1802.01746.*
- [33] Esposito, C., De Santis, A., Tortora, G., Chang, H., & Choo, K. K. R. (2018). Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?. *IEEE Cloud Computing*, 5(1), 31-37.
- [34] Yuan, Y., & Wang, F. Y. (2016, November). Towards blockchain-based intelligent transportation systems. In Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on (pp. 2663-2668). IEEE.
- [35] Cong, L. W., & He, Z. (2018). Blockchain disruption and smart contracts (No. w24399). National Bureau of Economic Research.
- [36] Nagothu, D., Xu, R., Nikouei, S. Y., & Chen, Y. (2018). A microservice-enabled architecture for smart surveillance using blockchain technology. *arXiv preprint arXiv:1807.07487.*
- [37] Angraal, S., Krumholz, H. M., & Schulz, W. L. (2017). Blockchain technology: applications in health care. *Circulation: Cardiovascular Quality and Outcomes*, 10(9), e003800.
- [38] Sharma, P. K., Chen, M. Y., & Park, J. H. (2018). A software defined fog node based distributed blockchain cloud architecture for IoT. *IEEE Access*, 6, 115-124.
- [39] Modum.io. modum.io — solution. Available at: <https://modum.io/system/>, 2018. Accessed on 29.01.2018.
- [40] Dentacoin.com. Dentacoin. Available at: <https://dentacoin.com/>, 2018. Accessed on 29.11.2018.
- [41] Tron.network. Tron. Available at: <https://tron.network/en.html>, 2018. Accessed on 29.11.2018.
- [42] Serguei Popov. The tangle. Available at: [https://iota.org/IOTA\\_Whitepaper.pdf](https://iota.org/IOTA_Whitepaper.pdf), 2017. Accessed on 23.11.2018.

- [43] Coinmarketcap.com. Cryptocurrency market capitalizations: sorted by trade volume. Available at: <https://coinmarketcap.com/>, 2018. Accessed on 15.11.2018.
- [44] Victor S. Adamchik. Concept of hashing. Available at: <https://www.cs.cmu.edu/~adamchik/15-121/lectures/Hashing/hashing.html>, 2009. Accessed on 25.11.2017.
- [45] Alt H. Dietzfelbinger M. Reischuk R. Scheideler C. Vollmer H. Wagner D Vcking, B. Algorithms unplugged. Available at: <https://link.springer.com/content/pdf/10.1007%2F978-3-642-15328-0.pdf>, 2011. Accessed on 25.11.2018.107
- [46] Douglas R. Stinson. Cryptography: Theory and practice, third edition. Available at: [http://www.ksom.res.in/files/RCCT-2014-III-CM/CryptographyTheoryandpractice\(3ed\).pdf](http://www.ksom.res.in/files/RCCT-2014-III-CM/CryptographyTheoryandpractice(3ed).pdf), 2014. Accessed on 04.11.2017.
- [47] Li, J., Greenwood, D., & Kassem, M. (2018, June). Blockchain in the built environment: *analysing current applications and developing an emergent framework*. Diamond Congress Ltd.
- [48] Blockchain, <https://www.forbes.com/sites/bernardmarr/2018/02/16/a-very-brief-history-of-blockchain-technology-everyone-should-read/#8c15937bc470> Accessed on 11-09-2018.
- [49] Blockchain Background, <https://www.dlapiper.com/en/denmark/insights/publications/2017/06/blockchain-in-background-challenges-legal-issues/> Accessed on 11-09-2018.
- [50] S. Nakamoto, *Bitcoin: A peer-to-peer electronic cash system* (2008).
- [51] B. Vitalik, The value of blockchain technology (2015. Last accessed november 2018). URL <https://blog.ethereum.org/2015/04/13/visions-part-1-the-value-of-blockchain-technology/>
- [52] Bitcoin developer guide (2017. Last accessed 17 November 2018). URL <https://bitcoin.org/en/developer-guide#block-chain>.
- [53] Bitcoin-Forum, Difference between miners and nodes (2016. Last accessed 21 July 2018). URL <https://bitcointalk.org/index.php?topic=1734235.0>
- [54] Ethereum computer built on embedded devices (2017. Last accessed 16 September 2018). URL <http://ethembedded.com/>
- [55] The-Linux-Foundation, Hyperledger business blockchain technologies (2018. Last accessed 15 September 2018). URL <https://www.hyperledger.org/projects>
- [56] P. Sebastia\_n, An introduction to ethereum and smart contracts: a programmable blockchain (2017. Last accessed 24 october 2018). URL <https://auth0.com/blog/an-introduction-to-ethereum-and-smart-contracts-part-2/>



- [57] D. Tuesta, J. Alonso, N. C\_amara, et al., Smart contracts: the ultimate automation of trust, *Digital Economy Outlook-October*.
- [58] S. Huckle, R. Bhattacharya, M. White, N. Belo\_, Internet of things, blockchain and shared economy applications, *Procedia Computer Science 98 (2016) 461{466}*.
- [59] A. Baliga, Understanding blockchain consensus models, *White paper, Persistent Systems Ltd. 2017*.
- [60] M. Vukoli\_c, The quest for scalable blockchain fabric: Proof-of-work vs. bft replication, in: *Proceedings of the International Workshop on Open Problems in Network Security, Springer, 2015, pp. 112{125}*.
- [61] Z. Zheng, S. Xie, H. Dai, X. Chen, H. Wang, An overview of blockchain technology: Architecture, consensus, and future trends, in: *Proceedings of the IEEE International Congress on Big Data (BigData Congress), 2017, pp. 557{564}*.
- [62] I. Eyal, E. G. Sirer, Majority is not enough: Bitcoin mining is vulnerable, *Commun. ACM 61 (7) (2018) 95{102}*. doi:10.1145/3212998. URL <http://doi.acm.org/10.1145/3212998>
- [63] A. Miller, Y. Xia, K. Croman, E. Shi, D. Song, The honey badger of bft protocols, in: *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, ACM, 2016, pp. 31{42}*.
- [64] N. Szabo, The idea of smart contracts, *IEEE International Workshop on Electronic Contracting (WEC)*.
- [65] N. Houy, It will cost you nothing to'kill'a proof-of-stake crypto-currency, *Econ.Bull 34 (2) (2014) 1038{1044}*.
- [66] EconoTimes, Blockchain project antshares explains reasons for choosing dbft over POW and pos (2017. Last accessed 18 october 2018). URL <http://www.econotimes.com/Blockchain-project-Antshares-explains-reasons-for-choosing-dBFT-over-PoW-and-PoS-659275>
- [67] Bitcoinwiki, Scalability (2017. Last accessed 19 September 2018).URL <https://en.bitcoin.it/wiki/Scalability>
- [68] D. Larimer, Delegated proof-of-stake (dpos), *Bitshare whitepaper*.
- [69] J. Kwon, Tendermint: *Consensus without mining, Draft v. 0.6, fall*.
- [70] Wüst, K., & Gervais, A. (2018, June). Do you need a Blockchain?. In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT) (pp. 45-54)*. IEEE.

- [71] Kern, A. G. (2018). *Blockchain Technology: a technology acceptance model (TAM) analysis (Doctoral dissertation)*.
- [72] Meng, W., Tischhauser, E. W., Wang, Q., Wang, Y., & Han, J. (2018). When intrusion detection meets blockchain technology: a review. *Ieee Access*, 6, 10179-10188.
- [73] Guegan, D. (2017). *Public Blockchain versus Private blockchain*.
- [74] Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017, June). An overview of blockchain technology: Architecture, consensus, and future trends. In *Big Data (BigData Congress), 2017 IEEE International Congress on* (pp. 557-564). IEEE.
- [75] <https://www.cio.com/article/3055847/security/what-is-blockchain-and-how-does-it-work.html> Accessed on 29-11-2018.
- [76] Jain, A., Jain, A., Chauhan, N., Singh, V., & Thakur, N. (2018). *Seguro Digital storage of documents using Blockchain*.
- [77] Xu, Q., Aung, K. M. M., Zhu, Y., & Yong, K. L. (2018). A blockchain-based storage system for data analytics in the internet of things. In *New Advances in the Internet of Things* (pp. 119-138). Springer, Cham.
- [78] Liang, G., Weller, S. R., Luo, F., Zhao, J., & Dong, Z. Y. (2018). Distributed blockchain-based data protection framework for modern power systems against cyber-attacks. *IEEE Transactions on Smart Grid*.
- [79] Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017, June). An overview of blockchain technology: Architecture, consensus, and future trends. In *Big Data (BigData Congress), 2017 IEEE International Congress on* (pp. 557-564). IEEE.
- [80] <https://en.wikipedia.org/wiki/integrity>, Last accessed 27-11-2018.
- [81] Sun, J., Yan, J., & Zhang, K. Z. (2016). Blockchain-based sharing services: What blockchain technology can contribute to smart cities? *Financial Innovation*, 2(1), 26.
- [82] Motro, A. (1989). Integrity= validity+ completeness. *ACM Transactions on Database Systems (TODS)*, 14(4), 480-502.
- [83] Wilkinson, S., Lowry, J., & Boshevski, T. (2014). Metadisk a blockchain-based decentralized file storage application. *Technical Report. Technical Report*.
- [84] Ethereum, W. G. (2014). A secure decentralised generalised transaction ledger [J]. *Ethereum project yellow paper*, 151, 1-32.
- [85] <https://en.wikipedia.org/wiki/Integrity> (Accessed 27-11-2018)
- [86] Sun, J., Yan, J., & Zhang, K. Z. (2016). Blockchain-based sharing services: What blockchain technology can contribute to smart cities. *Financial Innovation*, 2(1), 26.

- [87] <https://www.techopedia.com/definition/29059/data-ownership>, (Last accessed 27-11-2018).
- [88] <https://searchsecurity.techtarget.com/definition/access-control>, (Last accessed 27-11-2018).
- [89] Motro, A. (1989). Integrity= validity+ completeness. *ACM Transactions on Database Systems (TODS)*, 14(4), 480-502.
- [90] [https://www.google.com/search?client=firefox-ab&ei=Sxz9W6rxGZKvwOGg5m4Aw&q=what+is+synchronization&oq=what+is+synchronization&gs\\_l=psyab.1.0.0i67l2j0i7i30l8.757808.761893..763829...1.0..1.447.2374.1j5j2j2j1....3..0....1..gws-wiz.....0i71j0i8i7i30j0i13.MSB76eaDmPY](https://www.google.com/search?client=firefox-ab&ei=Sxz9W6rxGZKvwOGg5m4Aw&q=what+is+synchronization&oq=what+is+synchronization&gs_l=psyab.1.0.0i67l2j0i7i30l8.757808.761893..763829...1.0..1.447.2374.1j5j2j2j1....3..0....1..gws-wiz.....0i71j0i8i7i30j0i13.MSB76eaDmPY), (Last accessed 27-11-2018).
- [91] <https://www.weibull.com/hotwire/issue26/relbasics26.htm>, (Last accessed 27-11-2018).