Design of Access Control Framework for Big Data-as-a-Service (BDaaS) Platform

Santosh Kumar Sharma[1], Dr Ajay Pratap[2*] and Dr Harsh Dev[3]
[1&2] AIIT, Amity University Uttar Pradesh, Lucknow-226028
[3]Pranveer Singh Institute of Technology, Kanpur
** Address for correspondence*
AIIT, Amity University Uttar Pradesh, Lucknow-226028, Uttar Pradesh, India
Tel: +91-9936189152, Email: apratap@lko.amity.edu

Abstract

Big data-as-a-Service [BDaaS] platform is used widely to handle and process the large volume of data generated on daily basis from different source. Generated data is usually kept in cloud environment and there may be chances that the system will fail and become subject to numerous forms of assaults. Many researchers have been working to provide security and protection to the data available on cloud. Blockchain technology has emerged as such technology that is used to provide a secured, distributed, and decentralized environment in cloud environment. In this research paper, we are proposing a framework to control the access of data available on big data-as-a-service on cloud platform using blockchain technology as a tool. We are modifying the ciphertext policy- attribute based encryption (CP-ABE) technique and implementing on blockchain platform. Algorithms for all possible situations have been developed using combination of CP-ABE and blockchain for secured data access in BDaaS.

Keywords: *Big-Data-as-a-Service [BDaaS], Blockchain, Attribute Based Encryption, Access Control*

1. Introduction

Data has become very important and key asset for every organization because most of the decision making are based on these data. The data is generated from various sources such as data from sensor devices, social media data, data of educational organizations, and government data etc. These different types of data can be structured, semi-structured or unstructured in nature. Big Data-as-a-Service was introduced for handling and processing of these large volume of data [1]. BDaaS is a technique that combines the facility of storing data with computing capabilities of cloud computing environment wrapped with the processing power of big data. This model is useful for delivering the data, analyzing the data and database and processing platform, along with other service models like PaaS, SaaS and IaaS. Big Data-as-a-Service is known as cloud-based framework which provides end-to-end demand based big data solutions to the business organizations. It can also be understood as combined capabilities of Data as a Service, Hadoop as a Service and Data Analytics as a Service. Various service models can be chosen to fulfill the specific demands of users. Although there are a lot of benefits of using Big Data- as-a-Service platform but security and privacy of the data kept in this environment becomes very critical issue. Researchers have developed various methods, frameworks and architectures that can provide the security and privacy of data but still there is a scope to address the issue of security and privacy such as access control,

exposure of data, data breaches, and malicious adversary by cloud users [21][23]. Thus, we come to know that level of protection that is needed for big data security and privacy is not assured by the cloud providers.

It has been observed that, blockchain technology has become one of the good solutions for providing secure and decentralized environment for data [17][18]. Blockchain technology was introduced for exchanging digital currency, but it is useful in other application areas for providing security and privacy to the data such as Internet of Things (IoTs) [16], smart home [19], smart city, educational systems, and healthcare.

Bitcoin is one of the most well-known utilization of the blockchain tasks. Technically, Blockchain can be termed as distributed and decentralized blocks or ledgers that holds entire exchanges gathered in blocks that has finished at any point in the N/W. Blockchain is also popular Distributed Ledger Technology (DLT). Working of blockchain technology is based on Point-to-Point (P2P) network in which every node is required to maintain a copy of the blockchain ledger. Blockchain databases are not governed by any central regulatory authority in this system. This technology also ensures that
blockchain database is secured and protected from the various types of cyber-attacks.

Objective of using blockchain technology in BDaaS kept on cloud storage architecture is to provide more protected and secured environment for the users. Here we are not focusing on any specific resources in a single server, instead of that blockchain network distributes all of them among nodes [19]. Some researchers have worked on blockchain-based security mechanism, but most of them suggest a mechanism for such centralized server systems. We are the approach of decentralization concept of blockchain for security purposes. This is the area where a lot of work is to be done using the fusion of different techniques and methods.

We have proposed an access control framework and designed algorithm for secure access control of data kept on BDaaS in cloud platform. The proposed framework uses CP-ABE algorithm to provide secured data access. Access information is stored in blockchain in the form of smart contracts which are designed in the form of algorithm.

Contributions of the research paper are as follows:

1- Access control framework for Big Data-as-a-Service (BDaaS) has been proposed using decentralized and secure blockchain technology. Sequence of entire process is also presented in the paper through UML sequence diagram.

2- The proposed framework contains all the access policies in the smart contracts of blockchain network using customized form of ciphertext policy-attribute based encryption algorithm which is one of the very popular algorithms for access control. It also uses the digital signature for the authentication of data.

3- Algorithm for user key generation, attribute authority key generation, user key generation, encryption and decryption has been designed in this research paper.

Organization of the research paper can be understood as follows:

Section-2 presents the Background details in which we have discussed three major technologies used in this paper which are Big Data-as-a-Service (BDaaS) and Blockchain Technology. Literature Review is summarized in Section-3. Section-4 presents traditional access model, and its problems are identified. A framework for secured access control in BDaaS environment has been proposed in Section-5 which presents workflow, sequence diagram and access tree-based solution for the proposed problem. Section-6 presents all the algorithms required for secure access control for BDaaS environment using blockchain technology with CP-ABE algorithm. Analysis and Future work have been mentioned in Section-7 and finally, Section-8 presents the conclusion of this paper.

2. Background Details

This research paper is based on two core technologies, that are Big Data-as-a-Service (BDaaS) and cloud computing. Our data is available on BDaaS platform on cloud and its access is being controlled using blockchain technology. In this section, we are presenting foundation concepts of both BDaaS and blockchain useful for papers point of view.

2.1. Foundation of Big Data-as-a-Service

Big Data-as-a-Service can be considered as an umbrella term that is used for different services related to the management functions of big data and running in the cloud environment. BDaaS is in the analogy with the XaaS models that is based on Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Product as a Service (PaaS) model in cloud computing environment that is applicable to big data. BDaaS is a brand-new service category and addresses the variances in data structure and descriptions by encapsulating different data as a service. Now the users must take care about their needs and obtain the service whenever and wherever they want for data storage, data analysis and data visualization [25]. It has become very popular among users by helping them to increase the productivity and cut the expenses. BDaaS has three layers abstraction for the customers which are Big Data infrastructure as a Service, Big Data Platform as a Service and Big Data Analytics Software as a Service. Here big data infrastructure as a Service is combination of Storage as a Service and Computing as a Service, big data platform as a service is combination of data as a service and database as a service, and big data analytics is considered as software as a Service [24]. It involves some major advantages such as cost effectiveness, quick decision making, better data visualization, quick response, management of data, and data analytics. Many IT giants including as Amazon, IBM, EMC, Microsoft, Google, SAP and Oracle have inhabited the BDaaS market area and primarily focused for providing big data storage and analysis services. For instance, Greenplum is an EMC used for data storage and analysis that provides storage services and allow the users to use Hadoop for BDA. Independent BDA services are offered by Amazon Workspace and Microsoft offers BDA services through the Windows Azure Marketplace.

Security, Privacy and access method are one of the major concerns of users in BDaaS environment because it is a cloud-based services and the data is spread across many servers. As we already know that data manipulation, data theft, data loss, and

denial of services are very common security challenges in cloud computing platform, which are caused by due loss of control, and trust over data. Most of the cloud service providers do not guarantee for data security and access control. Therefore, it becomes very important for all the parties that are participating in BDaaS to reconsider the terms data security and access control. Big-data-as-a-Service is used for handling and processing of large amount of data that are generated from different sources. Collected data is stored in cloud computing platform. Therefore, it may suffer from single point of failure and, also it will be easy for attackers to target it. Many researchers are working in this domain to provide security to the system and protect the data that is kept on big data in cloud environment.

2.1.1. BDaaS Architecture

BDaaS is one of the new research domains that is combination of big data and cloud computing. It encapsulates the data as a service and shields the differences that may cause due to data structure and definitions. The users are just concern about what data or service they want to get and whenever and wherever to search, store, analyze and visualize. We are considering the architecture of BDaaS that is derived from service generated big data. Zheng et al. [9] has presented three major components of service generated big data that are logs, service QoS and service relationship. These components cause log visualization, performance diagnosis, fault tolerance, QoS prediction, service identification and service migration. These components are deriving Big Data Analytics Software-as-a-Service, Big Data Platform-as-a-Service and Big Data Infrastructure-as-a-Service. In this research paper, we are considering this architecture of BDaaS for secured access control as given in Figure 1.

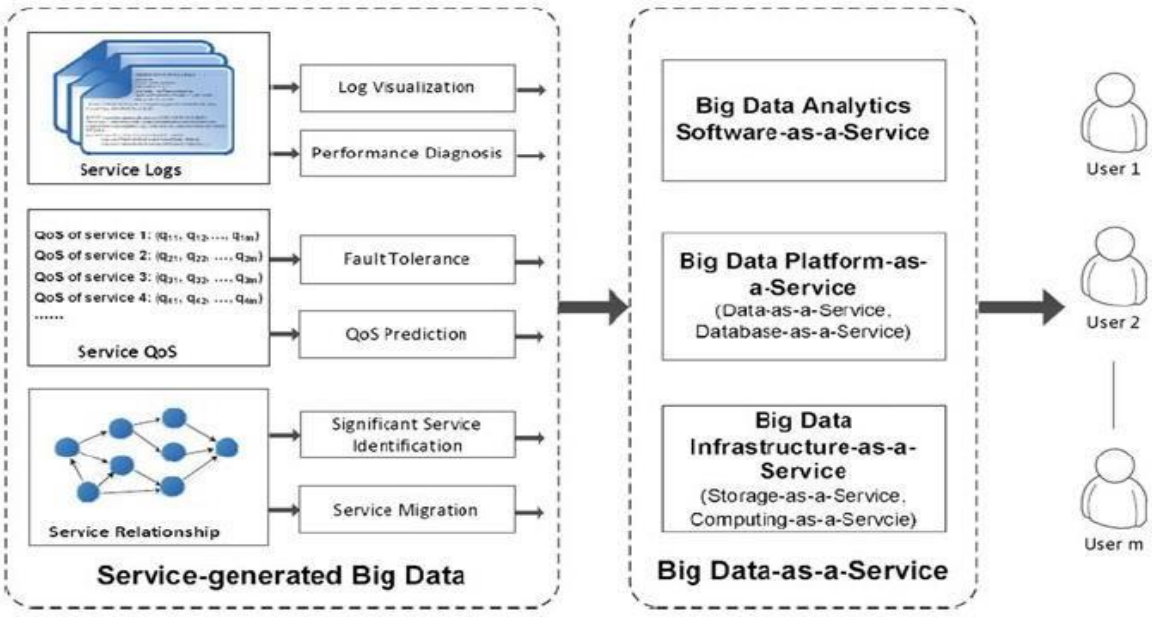

Figure 1. Architecture of Big Data-as-a-Service Platform

2.1.2. Comparison between Big Data and Big Data as-a Service

Big data refers to such data which is huge in size and increases with respect to time. It incorporates structured, unstructured, and semi-structured data. As we know

that due to its increasing size, we can't store it into traditional database management system. In this situation it needs specialized data management tool which can perform data storage, data analysis, data mining and also data visualization.

On other hand Big Data-as-a-Service (BDaaS) is a combination of software, large data Storage, infrastructure and a platform to deliver advanced data analysis on large data sets through cloud-based network. In other words, we can say that BDaaS is an umbrella term used for various services that are related to big data management functions running on cloud. This can be understood in analogy with XaaS models based on Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS). SaaS are still nonspecific but most of the user interactive services such as Email, CMS or CRM comes under this. Example of IaaS are virtual machines, networks, storage devices, or servers that is very basic building block and includes everything that is real or virtual, we would expect inside a data center. PaaS includes frequently employed software like web and database servers, or Hadoop and its ecosystem. This can be understood as a combination of data-as-a-service (DaaS), Hadoop-as-a-service (HDaaS), and data analytics-as-a-service.

Big data as a Service (BDaaS) = data-as-a-service (DaaS) + Hadoop-as-a-service (HDaaS) + data analytics-as-a-service (DAaaS).

Cloud computing is the backbone of BDaaS platform. Cloud computing refers to the on-demand availability of computing resources over internet. These computing resources can be servers, storage, databases, processing, software, analytics and networking. Cloud computing environment is very flexible and can be easily scaled as per the need and demand. Concept of cloud computing is based on three service models which are Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS). Analogy of XaaS model for BDaaS environment is also taken from these three service models of cloud computing. Amazon Web Service (AWS), Microsoft Azure, Google Cloud Platform, IBM Cloud Services are major vendors of cloud computing. These top cloud platform vendors offer big data technology support and services as Amazon EMR from Amazon Web Services (AWS), Google Cloud Dataproc and Microsoft's Azure HDInsight. Some other important big data as a service vendor are Cloudera, Databricks, HPE, Oracle and Qubole.

2.1.3. BDaaS Models and Its Challenges

We explored three different models that are being used in big data -as-a-service (BDaaS) environment. These models are very closely aligned with the three models of cloud infrastructure that are IaaS, PaaS, and SaaS. BDaaS models are as following:

1. Big Data Infrastructure as a Service (BDIaaS) – It include basic data services from a cloud service provider such as storage devices, virtual machines, and networks.
2. Big Data Platform as a Service (BDPaaS) – It offers frequently employed software those are provided by Amazon S3, EMR or RedShift. This doesn't include ETL processes and BI processes within it.

3. Big Data Software as a Service (SaaS) – It offers a complete big data stack within a single tool.

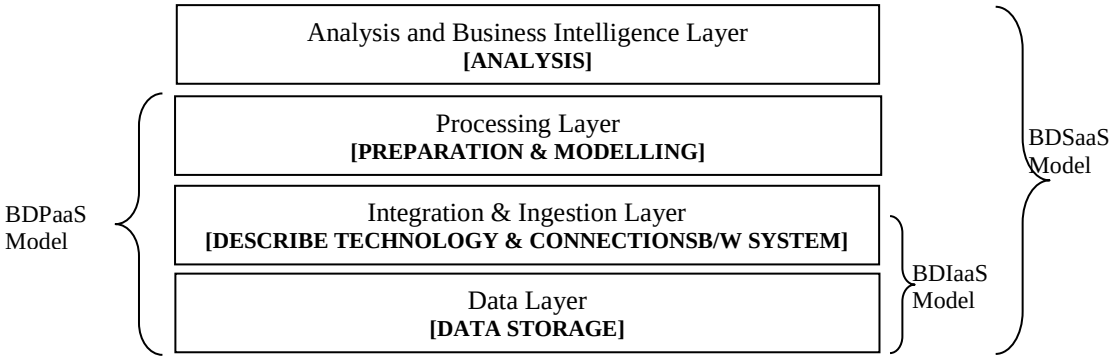

Figure 2. Formation of BDaaS Model

2.2. Concept of Blockchain Technology

The blockchain technology is among new rising decentralized and distributed technologies that have been developed in recent years. Application of blockchain in Bitcoin made it more popular. Blockchain is peer to peer database that is distributed in nature, and it maintains the continuously growing records (blocks) of transactions. These records (blocks) are linked with each other and secured by public key cryptographic technique.

From its name, we can see that blockchain consists of two parts as block and chain. Here blocks are storage units that store the records of transaction successfully completed in the system, and chain refers to connection between blocks of all time-stamped transaction blocks into continuous chain network [14]. In blockchain technology, new information is added to records (blocks) and they are distributed to all nodes participating in distributed system. Each block in the system is identified by its hash value using the secure hash cryptographic algorithm-25 bits (SHA-256) [16].

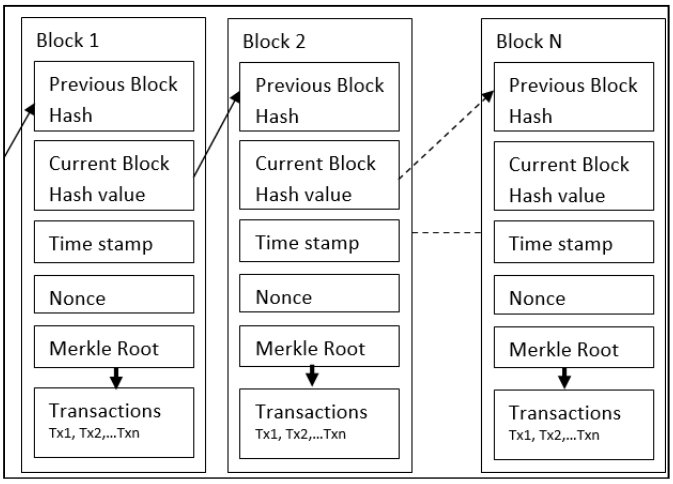

Figure 3. Basic Structure of Blocks

As given in figure 3. The hash value of parent block is stored and linked in the next continuous block. In case the content of any block is changed then secure hash value

will also be changed and this is sent to other nodes for validation. Each and every participant of blockchain network has its own private key to digitally sign and validate the transactions made by them.

2.3. Traditional Access Mechanism and Cryptography

Traditional model for access control was composed of four major components which were authentication of identity, access control, permission of access and audit if required. As shown in Figure 4, end user sends the request to access control mechanism. Identity authentication component of access control mechanism verifies the user's identity. After this, users can perform operations or access the services according to the access policy given in the authorization database.

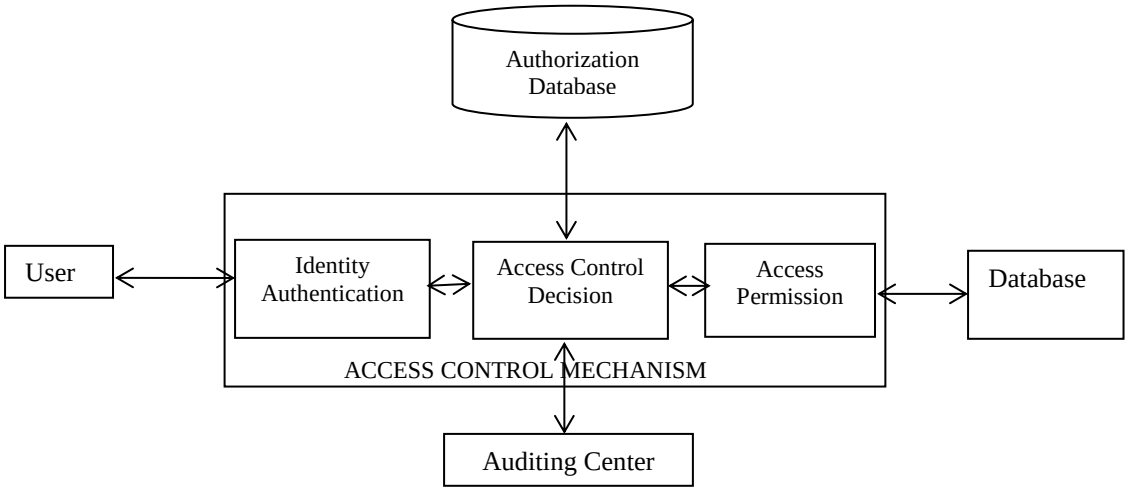

Figure 4. Traditional Access Control Technique

Traditional cryptographic techniques are not capable to improve the access control. The researchers started using these commonly used access control techniques which are Attribute Based Access Control (ABAC), History-Based Access Control (HBAC), Role Based Access Control (RBAC) and Capability Based Access Control (Cap-BAC).

3. Literature Review

Ciphertext-policy attribute-based encryption (CP-ABE) is a that is used to prevent users with other attributes from accessing data. The users with specific attributes can decrypt the encrypted data. This technique creates a secret key for users that is based on set of attributes. In this situation, decryption of the ciphertext is possible only when the attribute of the user's secret key matches with decryption policy. CP-ABE technique is used to control access that is proposed in IoT environment, and it also hides the access policy by using a hashing algorithm and provides security against insider attack using a signature verification scheme [1]. CP-ABE-based access control and revocation of services mechanism has been proposed on blockchain-based cloud storage system [2]. Encryption of data stored in the blockchain network by applying ciphertext-policy attribute-based encryption (CP-

ABE) and symmetric key algorithm is presented in [3]. Concept of smart offloading technique has been proposed by [4] that switches dynamically from full encryption to partial encryption according to a wise decision strategy based on ML algorithm considering the available resources and some encryption parameters such as number of attributes and the size of the data. A blockchain-based access control scheme (BacCPSS) for cyber physical social media (CPSS) big data is proposed for authorization, authorization revocation, access control and audit of real time data [5]. This article [6] has proposed building of an access control framework based on smart contract, built on the top of blockchain to secure the sharing of electronic medical records (EMR) in the smart healthcare system. Authors has performed review of application of blockchain technology for securing cloud storage [7]. Attribute-based encryption (ABE) is very powerful cryptographic approach for access control and fine-grained sharing on encrypted data. This functionality of ABE leads the adoption of ABE in encrypted cloud storage for flexible data sharing [9]. Blockchain-based Anonymous authentication with Selective revocation (BASS) for Smart industrial applications is proposed for smart industrial applications that support attribute privacy, selective revocation, and credential soundness [10]. Basic architecture of BDaaS has been presented in [11] that is considered for this research paper also. Random oracle model is used to handle security requirements such as mutual authentication and user anonymity and it also resists various malicious attacks [12]. Secure cloud storage framework with access control that combination of Ethereum blockchain and ciphertext-policy attribute-based encryption (CP-ABE) is has been proposed for secured access control [13]. [14] presents the review, opportunities and challenges of transforming big data using cloud computing resources. CP-ABE can enhance the security of access control on shared data with efficient authority verification [15]. Homomorphic encryption, order-preserving encryption schemes and Attribute-based encryption are described in this paper that provide data confidentiality and integrity [16].

Thus, we can see that blockchain technology attribute-based encryption are the technologies which are being used for various security aspects. Many authors have used different combination of technologies and concepts to handle different set of problems.

4. Problem Statement

Many researchers tried to solve the problem of access control using blockchain technology. Some of them are using ciphertext- policy attribute-based encryption for better access control and revocation. It has been observed that the combination of these two technologies can produce more better results. We have identified following problems after literature survey which are as followings:

(1) Owner of the data/ software/services available at BDaaS environment, are not able to decide about the people who can access the data at run-time.
(2) It is difficult to decide attributes for which access can be made possible for the users at the time of request.
(3) Dependency of access control architecture on semi-trusted authority.

5. Proposed Framework for secure BDaaS

Among these all-existing techniques, we are using attribute-based access control techniques. In this technique, access of data and services are granted or declined by evaluating a set of rules, policies, and relationships using the attributes of users, systems and environmental conditions. Ciphertext-Policy attribute-based encryption technique is used for secured access control on blockchain environment. Fusion of these two technologies are making more secured access control environment in the proposed work.

5.1. Ciphertext Policy- Attribute based Encryption [CP-ABE Algorithm]

Attribute based encryption (ABE) is very popular algorithm used for access control in which the secret key generated for user and the ciphertext depends upon the attributes. ABE comes under symmetric-key encryption and divided into two categories which are Key-policy attribute-based encryption (KP-ABE) and ciphertext-policy attribute-based encryption (CP-ABE).

In this research paper our access control algorithm is derived from CP-ABE algorithm which is implemented on blockchain platform. Access control is a way to limit the access of any system or resources in physically or virtually. We can also say that it is a security technique that has control over who can view or access anything. Figure 5 represents the general flow of work for CP-ABE algorithm.

Basic steps of CP-ABE algorithm are as following:

1- Generate Public Key and Master Key
2- Encrypt the message along with the access structure of all attributes, then final output will be the Ciphertext.
3- Generate Private key using Master key and the attributes used
4- Decrypt the message using Public Key and Private Key
5- If required, we can perform delegate step that will take the secret Key and return the Secret key for given set of attributes.

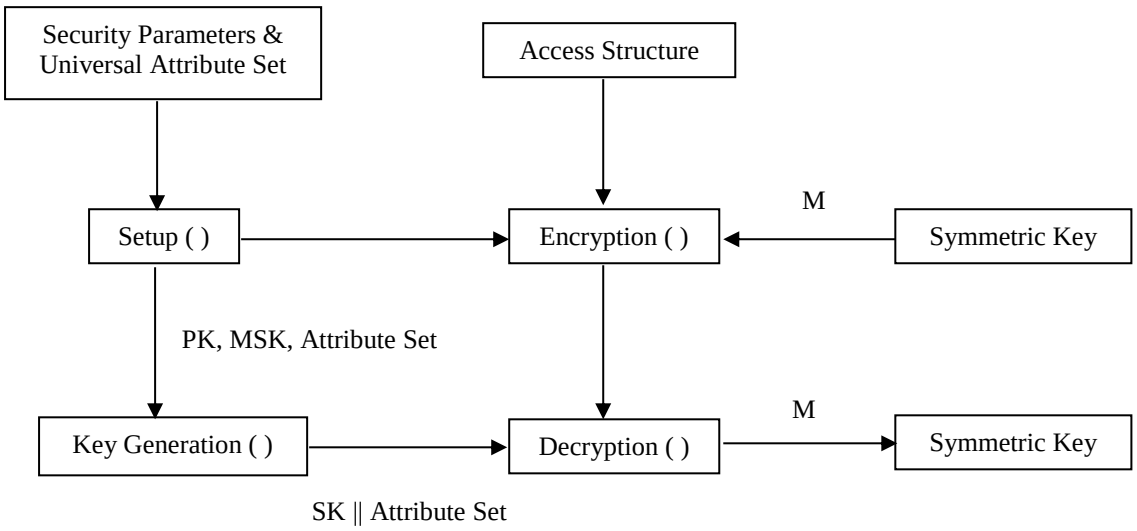

Figure 5. Flow diagram for CP-ABE Algorithm

5.2. Workflow of proposed Blockchain based Access Framework using CP-ABE

In the proposed model, workflow is the combination of three major technologies in which the BDaaS available on cloud is secured for access control using ciphertext policy- attribute based encryption (CP-ABE) algorithm and blockchain technology. Figure 6 presents the framework of proposed access control using blockchain and CP-ABE algorithm.

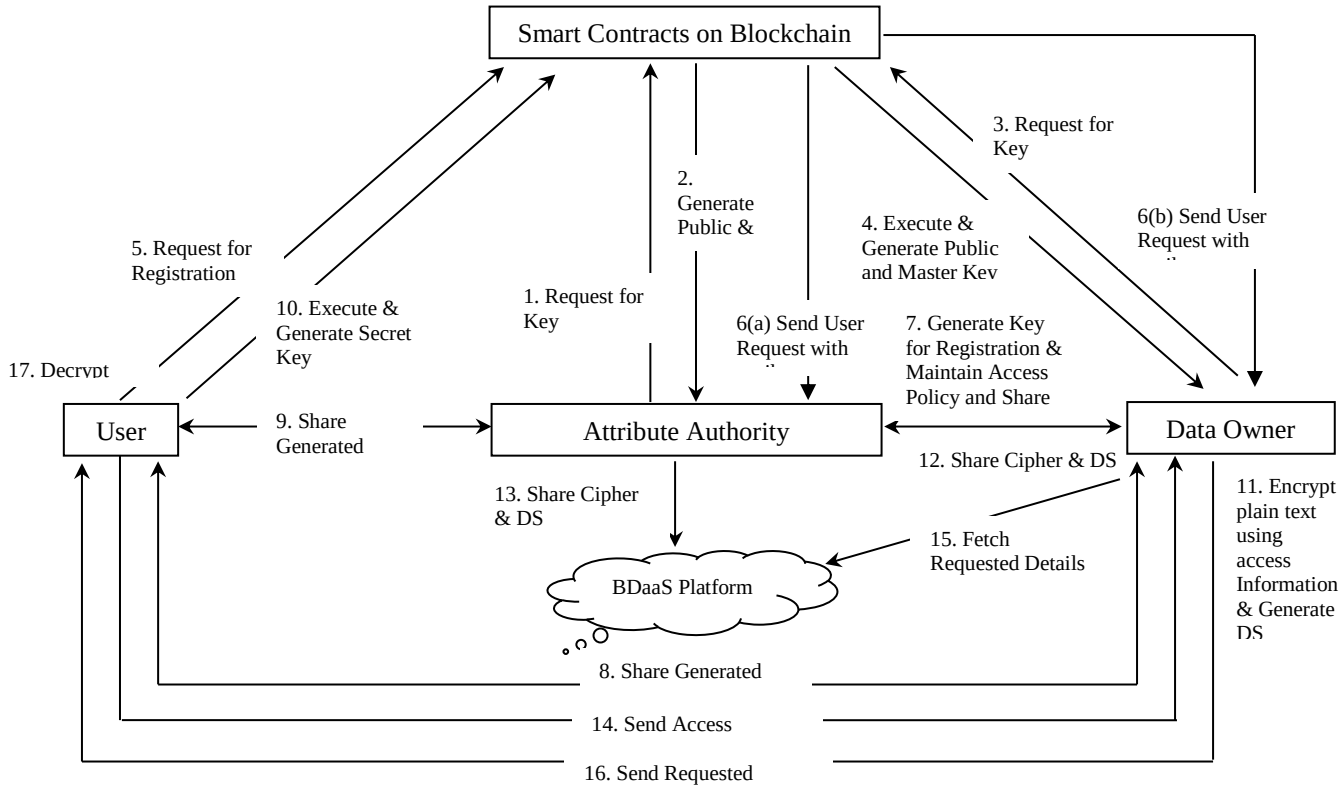

Figure 6. Proposed architecture for Blockchain based access control in BDaaS

Table 1. Representation Table

SN	Symbol	Description
1	GP	General Security Parameter
2	U	Universal Attribute Set
3	M	Data File
4	T	Access Set
5	DS	Digital Signature
6	CT	Ciphertext
7	S	Set of Attributes
8	$KeyGen_{(DO)}$	Key Generation Algorithm for Data Owner
9	$PK_{(DO)}$	Private Key of Data Owner
10	$MSK_{(DO)}$	Master Key of Data Owner
11	$KeyGen_{(AA)}$	Key Generation Algorithm for Attribute Authority

12	$PK_{(AA)}$	Private Key of Attribute Authority
13	$MSK_{(AA)}$	Master Key of Attribute Authority
14	$U_{(AT)}$	User Attribute
15	$SK_{(USER)}$	Secret Key of User
16	UserKeyGen ( )	Key Generation Algorithm for User
17	EncryptSign ( )	Algorithm for Encryption and Digital Signature
18	Decryptify ( )	Algorithm for Decryption and Verification

WORKFLOW FOR THE GIVEN FRAMEWORK

Attribute authorities and data owners send key generation requests to generate global parameters and public and master keys to register in the proposed architecture.

Step-1. Attribute authorities send request to blockchain network for key generation requests.

Step-2. Blockchain network generates public key [PK(AA)] and master keys [MSK(AA)] for registration in the proposed architecture and sends to Attribute Authorities.

Step-3. Service Owner send request to blockchain network for key generation.

Step-4. Blockchain network generates public key [PK(DO)] and master keys [MSK(DO]for registration in the proposed architecture and sends to Service Owner.

Step-5. User sends the request for registration to the blockchain network.

Step-6. Blockchain Network sends the details [U(AT)] (attributes such as ID, Name, Contact No, Email ID) of user to the Service Owner and Attribute Authority.

Step-7. The Service Owner and the attribute authority save the details of user and generates an access key using CP-ABE Algorithm corresponding to the user's list of attributes and shares with attribute authority.

Step-8. Service owners shares the generated access key [SK (DO, User)] with the user.

Step-9. Attribute authority shares the access key [SK (AA, User)] with the user and also generate access key for attribute group key [SK* (AA, User)].

Step-10. Now user generates secret key [SK(USER)] by executing key generation algorithm by using the keys generated by service owner and attribute authority.

Step-11. Service owner encrypts the plain text into ciphertext and digital signature.

Step-12. Service owner shares the cipher text with attribute authority.

Step-13. Cipher text generated in Step-11 is outsourced to Bigdata-as-a-Service Platform on cloud.

Step-14. Whenever the user sends the access request to data owner, then data owner takes it from Bigdata-as-a-Service Platform on cloud and shares with the user.

Step-15. The user performs decryption process on cipher text and verify the digital signature. User can do it only if they are authenticated user

5.3. Access Tree

Access policies in the proposed system will be represented in the form of tree. It uses AND, OR and K-Threshold gates for designing it. Leaf nodes represents the attributes for both negated and non-negated user details and can also use NOT gate.

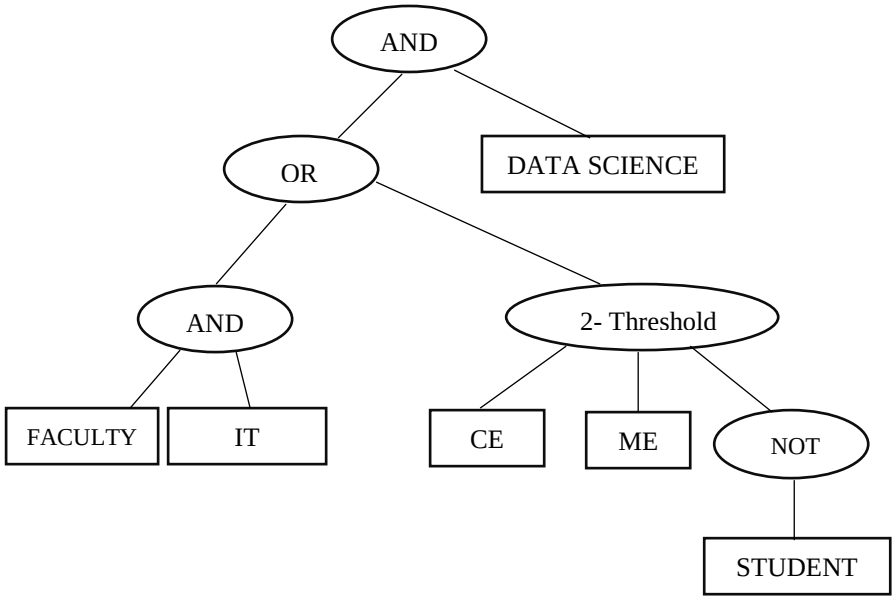

Figure 7. Tree-based Access Structure

Figure 7 represents the access structure of faculty who is given for access permission based on their attributes. Here leaf nodes are denoting non-negated attributes such as faculty, IT, ME, CE and data Science and one negated attribute student. As mentioned here attribute set {Faculty, IT, Data Science} which represents that the faculty members of IT department teaching data science are authorized for access. Whereas the second attribute set {Student, CE, ME} represents unauthorized access policy set means students of mechanical and civil engineering branch are not allowed to access the data.

5.4. UML Sequence Diagram for proposed Work

Sequence diagram is one of the very powerful and most commonly used way to show the interaction of system. This diagram depicts the interaction between various objects of system in a sequential order. Sequence diagram of proposed framework is presented here in Figure 8 for clear understanding of interaction among different components of system.

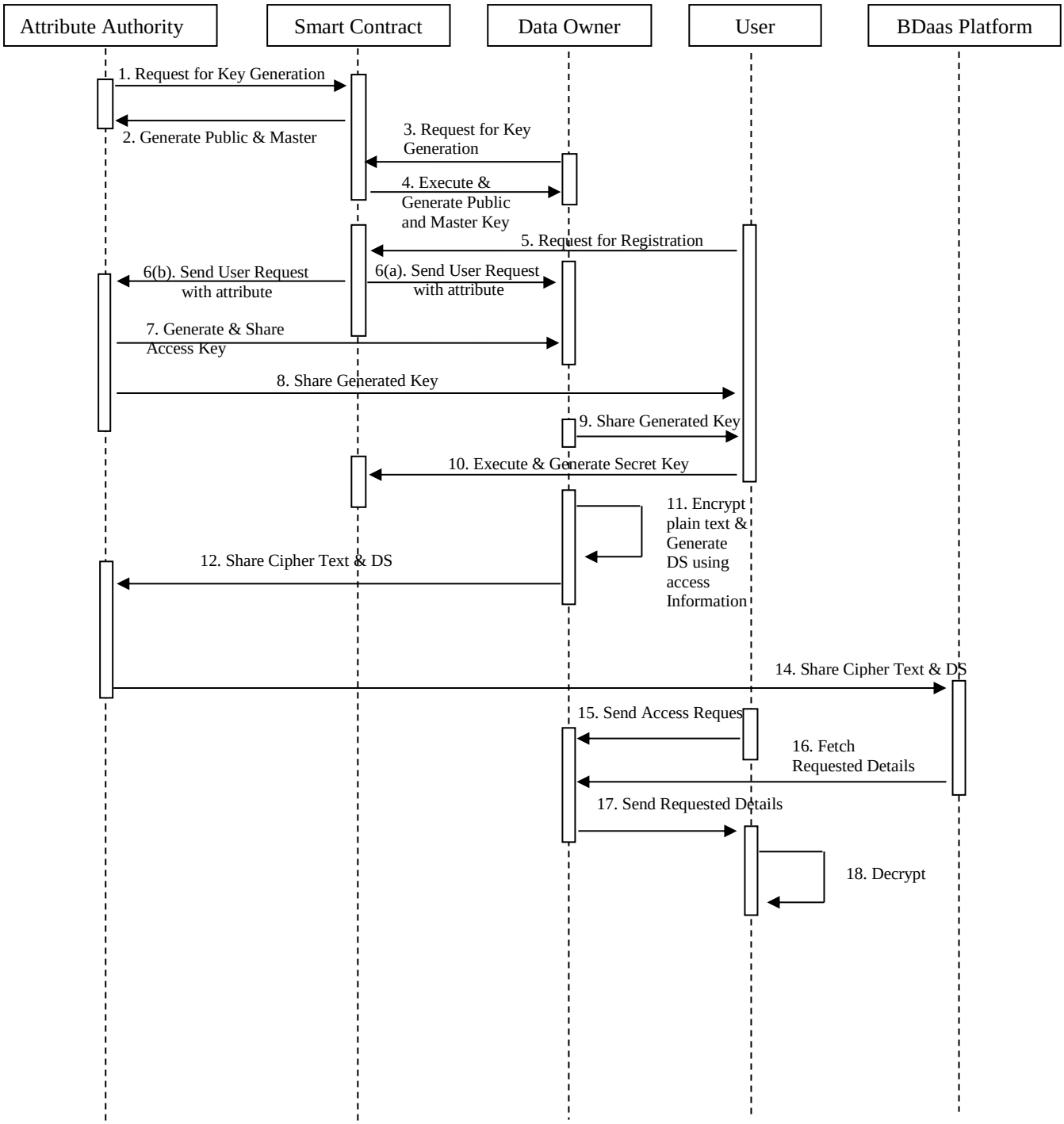

Figure 8. Tree-based Access Structure

6. Designing Smart Contract Algorithm for Proposed Framework

All the parameters of access control are handled by using smart contract which are simple programs and stored at blockchain. These programs execute when predefined condition given in smart contract are met. In the proposed model, entire access control is managed using different algorithms which will be written in ciphertext-policy attribute-based encryption. CP-ABE consists of four fundamental algorithms which are Setup, Encrypt, KeyGen and Decrypt.Setup algorithm takes implicit security parameters and provides public parameters public key (PK) and master key (MK). Encrypt algorithm takes

parameters as public key (PK), message (M) and access structure (A) which is represented as Encrypt (PK, M,A) and after encryption it provides cipher text (CT) which can be decrypted by only those users who possess attributes defined in access structure. Key generation algorithm takes master key (MK) and set of attributes (S) as a input which is represented as KeyGen (MK, S) and produces private key SK. After this the decrypt algorithm comes in the picture and takes public key (PK), cipher text (CT) with access policy and private key (SK) with set of attributes as an input and represented as Decrypt (PK,CT, SK) that produces message M after decryption if set of attributes satisfies the access structure.

By analyzing the workflow of proposed framework, we are identifying all smart contract algorithm used for access control, which are as following:

1. Public Key and Master Key Generation for Registration Process
2. User Key Generation Algorithm
3. Algorithm for Encryption
4. Algorithm for Re-encryption
5. Algorithm for Decryption

All smart contracts functions will be designed using CP-ABE algorithm and data owner and attribute authorities are involved to provide all required services to the users.

6.1. Algorithm for Setting Security Parameters

Setup (GP, U) takes the general security parameters and universal attribute set and generates the private key and master key for the owner and attribute authority. $KeyGen_{(DO)}$ (GP) and $KeyGen_{(AA)}$ (GP,c,d) are the two algorithms that are the parts of this setup.

6.2. Key Generation Algorithm for Data Owner

Input:
General Security Parameter (GP)
Output:
Private Key of Owner [$PK_{(DO)}$]
Master Key of Data Owner [$MSK_{(DO)}$]
Algorithm:
This algorithm will be denoted as
$KeyGen_{(DO)}$ (GP) $\rightarrow \{PK_{(DO)}, MSK_{(DO)}\}$

Step 1-	Begin
Step 2-	Choose a random number ‘a’ from the finite field over prime number p. $a \in Zp$
Step 3-	Choose a generator ‘g’ which fulfill the criteria: $b \leftarrow ga$ Here b is also from the finite field over prime number p. i.e. $b \in Zp$
Step 4-	$PK(DO) \leftarrow b$
Step 5-	$MSK(DO) \leftarrow a$
Step 6-	End

6.3. Key Generation Algorithm for Attribute Authority

Input:

General Security Parameter (GP)

Random Numbers c and d from Zp*

Output:

Private Key of Attribute Authority [$PK_{(AA)}$]

Master Key of Attribute Authority [$MSK_{(AA)}$]

Algorithm:

This algorithm will be denoted as

$KeyGen_{(AA)}$ (GP,c,d) $\rightarrow \{PK_{(AA)}, MSK_{(AA)}\}$

Step 1-	Begin
Step 2-	Choose a random number c from Zp*.
Step 3-	Choose a random number d from Zp*.
Step 4-	$PK_{(AA)} \leftarrow e(g,g)c$
Step 5-	$MSK_{(AA)} \leftarrow gc$
Step 6-	$PK^*_{(AA)} \leftarrow gd$
Step 7-	End

6.4. Key Generation Algorithm for Users

Attribute authority and data owners receive the attributes of user and generate access keys. These access keys are used to generate a secret key for the user. We can use some such protocol by which they can generate a secret key without telling their own keys. Two party computation function can be used to generate such key.

Algorithm: UserKeyGen ($MSK_{(DO)}$, $MSK_{(AA)}$, $U_{(AT)}$) $\rightarrow SK_{(USER)}$

Step-1.	Begin
Step-2.	Attribute Authority and Data Owner authenticate the User
Step-3.	Data Owner selects random exponent which is unique and secret to user i.e. $m \in Z^*_P$
Step-4.	Attribute authority selects random exponent i.e. $n \in Z^*_p$
Step-5.	Define two party computation function as $F = 2PC(DO(m,a), AA(c)) = (c + m)\, a$
Step-6.	Data Owner computes the parameter M $M = g^{(F/n)} = g^{(c + m)\, a / n}$
Step-7.	Attribute Authority compute the parameter N $N = M^{(1/a2)} = g^{(c + m)\ / n\, a}$
Step-8.	Data Owner sent parament M to Attribute authority and attribute authority sent the parameter N to data owner using 2PC for computation of secret key. a. Attribute Authority generate secret key $SK_{(AA, User)} = N^n = g^{(c + m)\ /\ a}$ b. Data Owner generates secret key $SK_{(DO, User)} = (D_i = g^n \cdot H(i)^{Ri}, D_i = g^{Ri})$ For all $i \in$ Tree and $Ri \in Z^*_P$ c. Attribute authority generates another secret key for attribute group key $SK^*_{(AA, User)} = H(User)^C$
Step-9.	Secret keys of data owner and attribute authority is used to generate $SK_{(USER)} = \{SK_{(AA, User)}, SK_{(DO, User)}\}$

Input:
Master Key $MSK_{(DO)}$ and $MSK_{(AA)}$
User Attribute $U_{(AT)}$
Output:
User Secret Key $[SK_{(USER)}]$

6.5. Algorithm for Data Encryption and Digital Signature

This algorithm is executed by data owner to encrypt the data according to the given access tree structure. After the encryption of data, data owner sent it to the big data platform available on cloud. This algorithm requires three parameters which are public key, data to be encrypted and access structure. Here public key of two parties i.e. data owner and attribute authority are involved so, both public key of data owner $PK_{[DO]}$ and $PK_{[AA]}$ will be used for encryption.

Input:
Public Key of Data Owner $[PK_{(DO)}]$,
Public Key of Attribute Authority $[PK_{(AA)}]$,
Plain Text/ Data [M],
Access Tree Structure [ATS]
Output:
Digital Signature [DS]
Cipher Text [CT]
Algorithm: EncryptSign $(PK_{(DO)}, PK_{(AA)}, M, ATS) \rightarrow DS, CT$

Step 1-	Start
Step 2-	For each node x in access tree ATS Choose a polynomial P_x
Step 3-	If node is Root Node then
	Set $P_x(0) = s$
Step 4-	else
Step 5-	If x is any other point of polyniomial
Step 6-	Set $P_x(0) = P_{parent(x)}(index(x))$
Step 7-	Calculate parameter $C' = M \cdot e(g.g)^{as}$
Step 8-	Calculate parameter $C = h^s$
Step 9-	else
Step 10-	If node y is leaf node then
Step 11-	Calculate parameter $C_Y' = g^{Py(0)}$
Step 12-	Calculate parameter $C_Y = H(attr(Y)^{Py(0)})$
Step 13-	Calculation of Final Cipher Text is done as:
Step 14-	$CT = [ATS, Sign = h(M), C' = M \cdot e(g,g)^{as}, C = h^s, (C_Y = H(attr(Y)^{Py(0)}), C_Y' = g^{Py(0)}$ for all x)]

6.6. Algorithm for Data Decryption and Verification

Input:
Secret Key of User $[SK_{(USER)}]$,
Cipher Text [CT],
Digital Signature [DS]
Output:

Digital Signature Verification [Success/ Failure]
Plain Text / Data [M]
Algorithm: Decryptify ($SK_{[USER]}$, $PK_{[AA]}$, CT) → Success/ Failure, M

Step 1-	Start
Step 2-	When x is leaf node &&
Step 3-	if i ∈ S and i = order(x) then
Step 4-	Decryptify ($SK_{[USER]}$, $PK_{[AA]}$, CT, x) = $e(D_j, C_y) / e(D'_i, C'_y)$
Step 5-	If i S the
Step 6-	Decryptify ($SK_{[USER]}$, $PK_{[AA]}$, CT, x) = Null
Step 7-	When policy is satisfied by access tree then:
Step 8-	Decryptify ($SK_{[USER]}$, CT, z) = $e(g, g)^{FS}$
Step 9-	$C' = F. e(g, g)^{aS}$, $e(C,D) = e(g^{aS}, g^{a+n/b})$
Step 10-	$M = C'/ e(C,D)/S$
Step 11-	Verify (M, DS, PUK)
Step 12-	Compute $e(h(M), g^x)$
Step 13-	Stop

8. Conclusion

As we know that, privacy and security of data kept on BDaaS platform are challenging issues. The proposed framework, we have designed the framework for secured data access in BDaaS platform using blockchain technology in combination with CP-ABE algorithm. Algorithm for secured access control has been designed according to blockchain environment. Setup phase generates two algorithms which are for private key and master key generation of both data owner and attribute authority. Once the user registers on blockchain network, data owner and attribute authority generate secret key for that user. Both secret keys are passed through two-party computation function to generate one common key for the user. Now the data encryption is performed, and digital signature of data is generated. The user decrypts the data using its secret key and also verify the data.

It has been planned to present a case study of this new framework and compare this new designed framework with some existing ones.

References

[1] *P. Chinnasamy,, P. Deepalakshmi, A. K. Dutta, J. You, and G. P. Joshi, "Ciphertext-Policy Attribute-Based Encryption for Cloud Storage: Toward Data Privacy and Authentication in AI-Enabled IoT System", vol. 10, no. 68, Mathematics* ***(2022)****, https://doi.org/10.3390/math10010068.*

[2] P. Sharma, R. Jindal, M. D. Borah, "Blockchain-based cloud storage system with CP-ABE-based access control and revocation process" The Journal of Supercomputing. ***(2022)****, DOI: 10.1007/s11227-021-04179-4.*

[3] *T. Lee, H. Moon, J. Jang, "Data encryption method using CP-ABE with symmetric key algorithm in blockchain network", International Conference on Information and Communication Technology Convergence (ICTC).* ***(2021)****, DOI: 10.1109/ICTC52510.2021.9620889*

[4] *M. B. Taha, H. O. Slimane, C. Talhi, "Smart offloading technique for CP-ABE encryption schemes in constrained devices", SN Appl. Sci. 2, no. 274.* ***(2020)****, https://doi.org/10.1007/s42452-020-2074-z*

[5] *T. Liang, N. Shi, C. Yang and K.Yu, "A Blockchain-Based Access Control Framework for Cyber-Physical-Social System Big Data", IEEE Access, vol.4* ***(2020)*** *DOI: 10.1109/ACCESS.2020. 2988951*

[6] *Q. Saini, N. Zhu, Y. Singh, L G Xiang, Y. Zhang, "A smart-contract-based access control framework for cloud smart healthcare system". IEEE Internet Things J 8(7):5914–5925,* ***(2020)****. https:// doi. org/ 10. 1109/ JIOT. 2020. 30329 97.*

[7] *P. Sharma, R. Jindal, MD. Borah, "Blockchain technology for cloud storage: a systematic literature review", ACM Computer Survey vol. 53, no. 4,* ***(2020)****.*

[8] *Q. Su, R. Zhang, R. Xue, P. Li, "Revocable attribute-based signature for blockchain-based healthcare system", IEEE Access 8:127884–127896,* ***(2020)****. https:// doi. org/ 10. 1109/ ACCESS. 2020. 30076 91*

[9] *H. Zheng, J. Shao, G. Wei, "Attribute-based encryption with outsourced decryption in blockchain", Peer-to-Peer Network Application,* ***(2020)****, pp. 1643–1655.*

[10] *Y. Yu, Y. Zhao, Y. Li, X. Du, L. Wang, M. Guizani, "Blockchain-based anonymous authentication with selective revocation for smart industrial applications", IEEE Trans Ind Inf, vol. 16, no. 5,* ***(2020)****, pp. 3290–*3300*. https:// doi. org/ 10. 1109/ TII. 2019. 29446 78.*

[11] *S.K. Sharma, A. Pratap, H. Dev, "Challenges against Big Data as a Service: A Survey", International Journal of Computer Science and Engineering, vol.7, no.12, (2019), pp. 74-78.*

[12] *L Xiong, F. Li, S. Zeng, T. Peng Z. Liu, "A blockchain-based privacy-awareness authentication scheme with efficient revocation for multi-server architectures", IEEE Access 7:125840–12585,* ***(2019)****, https:// doi. org/ 10. 1109/ ACCESS. 2019. 29393 68.*

[13] *S. Wang, X. Wang, Y. Zhang, "A secure cloud storage framework with access control based on blockchain" IEEE Access 7:112713–112725,* ***(2019)****. https:// doi. org/ 10. 1109/ ACCESS. 2019. 29292 05*

[14] *M. Muniswamaiah,T. Agerwala,and and C. Tappert,"Big Data in Cloud Computing Review and Opportunities", International Journal of Computer Science & Information Technology (IJCSIT), vol. 11, no 4,* ***(2019)****, pp. 43-57.*

[15] *L. Zhang,Y. Cui, Y. Mu, "Improving security and privacy attribute-based data sharing in cloud computing", IEEE Systems Journal, vol. 14, no. 1,****(2019)****, pp. 387-397.*

[16] *H. Shekhawat, S. Sharma, R. Koli, "Privacy-Preserving Techniques for Big Data Analysis in Cloud", Second International Conference on Advanced Computational and Communication Paradigms,* ***(2019)****, pp. 1-6.*

[17] *N. Elisa, L. Yang, F. Chao, Y. Cao, "A framework of Blockchain-based secure and privacy-preserving E-government system", Wireless Networks,* ***(2019)****, pp. 1-11.*

[18] X. Wang, L.T. Yang, L. Huazhang, MJ.Deenu, "A Big Data-as-a-Service Framework: State-of-the-Art and Perspectives", IEEE Transactions on Big Data, vol. 4, no. 3, ***(2018)****, pp. 325-340.*

[19] S. Khan, K.A. Shakil, S.A. Ali, M. Alam, "On designing a generic framework for big data-as-a-service", 1st International conference on advanced research in engineering sciences (ARES), ***(2018)****, pp. 1-5.*

[20] H. ES-Samaali, A. Outchakoucht and J. P. Leroy, "A Blockchain-based Access Control for Big Data", International Journal of Computer Networks and Communications Security, vol. 5, no. 7, ***(2017)****, pp. 137–147.*

[21] L. Yue, H. Junqin, Q. Shengzhi, W. Ruijin, "Big data model of security sharing based on Blockchain", 3rd International Conference on Big Data Computing and Communications, ***(2017)****, pp. 117-121.*

[22] A. A. Claudio, C. Paolo, D. Ernesto, "Big Data Analytics as-a-Service: Issues and challenges", IEEE International Conference on Big Data, ***(2016)****, pp. 3638-3644.*

[23] A. Rahmani, A. Amine and M. R. Hamou, "A mathematical model of access Control in big data using Confidence interval and digital Signature", Fourth International Conference on Advanced Information Technologies and Applications, ***(2015)****, DOI: 10.5121/csit.2015.51515*

[24] Z. Zheng, J. Zhu, M.R. Lyu, "Service-generated big data and big data-as-a-service: an overview", IEEE international congress on Big Data, ***(2013)****, pp. 403-410.*

[25] E. Xinhua, H. Jing, W. Yasong, L. Lianru, "Big Data-as-a-Service: Definition and architecture", ICCT2013, ***(2013)****, pp. 738-742.*