OPTIMIZATION OF THE BIG DATA SECURITY BASED ON THE COMBINATION OF IMPROVED ATTRIBUTE ENCRYPTION AND THE BLOCKCHAIN.

RAVELOJAONA Andriantsitoherintsoa,
PHD Student, ED STII, University of Antananarivo, PB 101 *Madagascar*
andojaonaa@gmail.com

RAKOTOMANANA René,
Doctor in cognitive Science and application, ED STII, University of Antananarivo, PB 101 *Madagascar*
reneheli@yahoo.fr

ANDRIAMANOHISOA Hery Zo,
University of Antananarivo, Madagascar
Professor, ED STII, University of Antananarivo, PB 101 *Madagascar*
aheryzo@gmail.com

July 29th 2024

Abstract

This article shows a new way of combining the ciphering through improved attributes and the Blockchain for the optimal safety in Big Data.

After encrypting a chain of hashed blocks, ordered, treated safely in the Edge/Fog nodes, they are stocked safely on the cloud Service Center. The authorized users are the only ones that can have access to them.

Keys Words: *CP-ABE, elliptic curves, Cloud, Big Data, Access, security, optimization, Blockchain.*

1.INTRODUCTION

The Security of the Data are generally focused on three objectives: confidentiality, integrity and availability. The confidentiality consists of protecting the Data from non-authorized access. The integrity aims at the protection from non-authorized changes. The availability assures the accessibility to Data for authorized users. [1]

The realization of these objectives of security in the field of Big Data is often compared with different challenges imposed by the need of shared and upgradable architectures, the heterogeneousness of the environments, the sharing out of resources with the foreign systems.

Mots clefs : CP-ABE, elliptique courbe, Cloud, Big Data, Accès, sécurité, optimisation,

2.PROPOSED SOLUTION

2.1 Mathematical modelization of Big Data

The numerical data with varied shapes, be it a picture, biochemical measures or marketing data are shown by a vector $x^i \in R^d$, where d the dimension, is often superior to 10^6.

$$x^{i}= (x_1, x_2, x_3, \dots, x_k, x_{k+1}, \dots, x_{n-1}, x_n) \quad (1)$$

Where x_i are kinds of the numerical data which are parts of the numerical data x^{i}, the exposant i indicates the number of the considered data, $i \in \mathrm{N}$.

The big Data is the bulky unity data [2], we are going to modelize mathematically with a vector X.

$$X = (x^1, x^2, x^3, \dots, x^{p-1}, x^{p}, x^{p+1}, \dots, x^{q}) \quad (2)$$

2.2 System structural

The figure 1 shows the global architecture of our system which is composed of:
. A producer or Data Source (DS), responsible for producing and ciphering the Data.
. A Data consumer (RQ) which receives and deciphers them.
. An authority of trust responsible for keeping the keys and attributes.
. An entity of externalization of certain treatments, notified (PR) for proxy. In a true situation, these entities will be taken for example by the technologies of Fog computing.

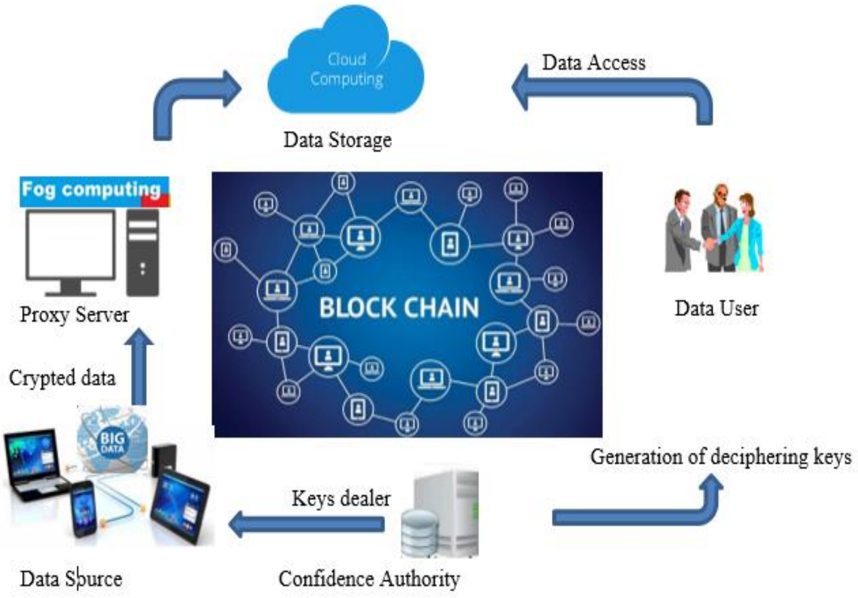


Fig- 1: System architectural

2.3 Ciphering through attribute and blockchain.

In order to assure the security of our proposition, the administration of the access control of messages can be done by Blockchain, a decentralized solution allowing to overcome the tough problem of unicity of the failure point (in the field of confidence).

2.3.1 Ciphering and deciphering through attribute

Our algorithm can be divided into 3 parties quoted as follows:

Preparation in which every Data consumer cons, gets an identity from AC which defines the kinds of the public parameters and the main keys of the system.

Ciphering of messages M are ciphered with the help of the symmetrical keys generated by ECDH protocol [3],[4] .

Deciphering of messages can be done after getting the keys calculated by AC, and delivered to the authorized access according to the Blockchain process [5].

Algorithm 2.01Ciphering algorithm through attributes

Data in entrance : $M, P_K, \Gamma, (A, B), E(a, b), G;$
Data in exit : C

```
Variables : $\alpha_x, \beta_x, \mu, \nu$
Beginning
  $s \in [1 \quad n-1]$, // uncertain Choice of a number s;
  $C0 \leftarrow M + s\,G$ ;// Initialization of value $C_0$
  $\mu \in [1 \quad n-1]^l$, // uncertain choice of a vector $\mu$;
  $\beta_x \leftarrow A_x \cdot \mu$ ;//Calculation of the value $\beta_x$
  $\nu \in [1 \quad n-1]^l$, // uncertain choice of a vector $\nu$;
  $\alpha_x \leftarrow A_x \cdot \nu$ ;//Calculation of the value $\alpha_x$
  $C_{1,x} \leftarrow \alpha_x \cdot G + \beta_x \cdot Pk_{R(x)}$
  $C_{2,x} \leftarrow \beta_x \cdot G$ ;//Calculation of $C_{i,x}$
end
return
```

Algorithm 2.02: Generation of keys algorithm

```
Data in entrance : $MSK$ ; $id$ ;
Data in exit : $CD_i$
Variables : i, $k_i$
Beginning
  for $i = 1$ to n to do
    $k_i \in [1 \quad n-1]$, // choice of a whole $k_i$;
    $CDi, id \leftarrow k_i + H(id).MSK$ ;// Calculation of ciphering keys $CD_{i,id}$ ;
End
Return $CD_{i,id}$
```

Algorithm 2.03: deciphering algorithm through

```
Data in entrance : $C, CD_{i,id}$
Data in exit : $M$
Variables : i, $k_i$
beginning
  $A_x \in A$, // Choice of $A_x$ , $(1,0,0, \ldots..0) \in A_x$ ;
  $C_1 \leftarrow \sum C_{1,x}$,  $\forall x$ // Calculation of $C_1$
  $C_2 \leftarrow \sum C_{2,x}.CD_{R(x),id}$,  $\forall x$ // Calculation of $C_2$
  $c_x \in [1 \quad n-1]$, // Choice of whole $c_x$;
  $s\,G \leftarrow \sum c_x\,(C_1 - C_2)$, // Calculation of $s\,G$
  $M \leftarrow C_0 - s\,G$, // Calculation of $M$
End
return $M$
```

2.3.2 Blockchain

To manage fully the messages of the access control, the paradigm BLOCKCHAIN can be used as dealt database. The recording of the access authorization on Blockchain can be done in token form, called « token ».

Algorithm 2.04 : Blockchain algorithm process

```
Data in entrance : $C, H$
Data in exit : $CD_{i,idx}$
Variables : i
Beginning
    $idx \leftarrow H(C)$// proxy calculates $idx$ ;
    Token($idx, @CLD, @Pr, @DO$)// sends from $idx$ to Data owner DO ;
    (Token($idx, @CLD, @Pr, @DO$), $@DO, @CLD$)//
        Storage of Data ciphered in the cloud;
    Si $a_i \in A$ does // verification of access permission,
        (Token($CD_{i,id}, @DO$), $@DO, @Cons_i$)// Validation of Data access ;
        (Token($idx, @CLD, @Pr, @DO$), $@CLD, @Cons_i$) //Data access,
    End if
End
Return $CD_{i,idx}$
```

2.4 Analysis and results

2.4.1 CP – ABE Security

The strengthening of the ciphering security CP-ABE we have proposed is fullfilled by the Diffie-Hellman protocol combination and the elliptical curve (ECDH). The enemy that wants to intercept the communication between the outside data and the proxy tie must solve the problem of discreet logarithm on this curve which is more difficult to solve than the factorization of entire [6]. It is quite impossible to hide the symmetrical key *MSK*. Moreover, the comparison by collusion among consumers are ineffectual, because we have generated an uncertain value to randomize the private key of every consumer.

According to our architecture, the Cloud and the proxy tie cannot recover the flat message. Therefore, the cloud is not included in the ciphering process/ deciphering of Data. The authorized Data consumer are the only to be allowed to use the ciphered, stored in the Cloud.

2.4.2 Blockchain Security

In order to assure the independent validation of the access control events, our system is conducted by the BLOCKCHAIN paradigm. The use of pseudonym in the blockchain transactions helps to keep the users' confidentiality.

Thanks to the numerically signed transactions, an enemy cannot falsify a control message or borrow the identity of the legitimate user. Therefore, the system ties are protected from the enemies who want to compromise them. This architecture provides security because each transaction requires a transaction signature.

2.4.3 Evaluation of the CP-ABE performance's

The performances of our system are evaluated according to the number of the connected attributes which is an important parameter worth being taken, considering its influence on the time of ciphering and of the time of deciphering.

We simulate the algorithm of the diagram origin ciphering CP-ABE [7] with the number of the defined attributes from.

N = {50, 100,150,200,250,300,350,400,450, 500}.

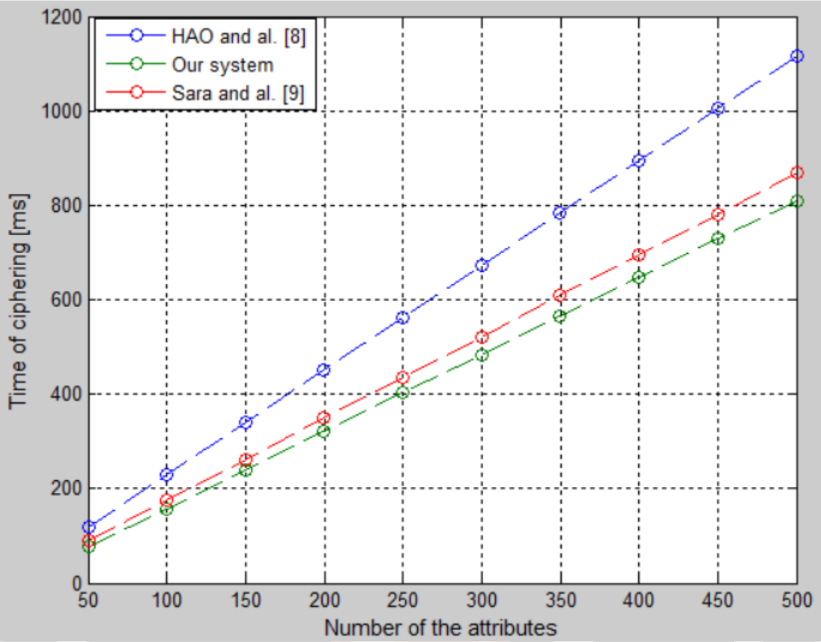


Fig- 2 : comparison of the ciphering performances of our system with other protocol

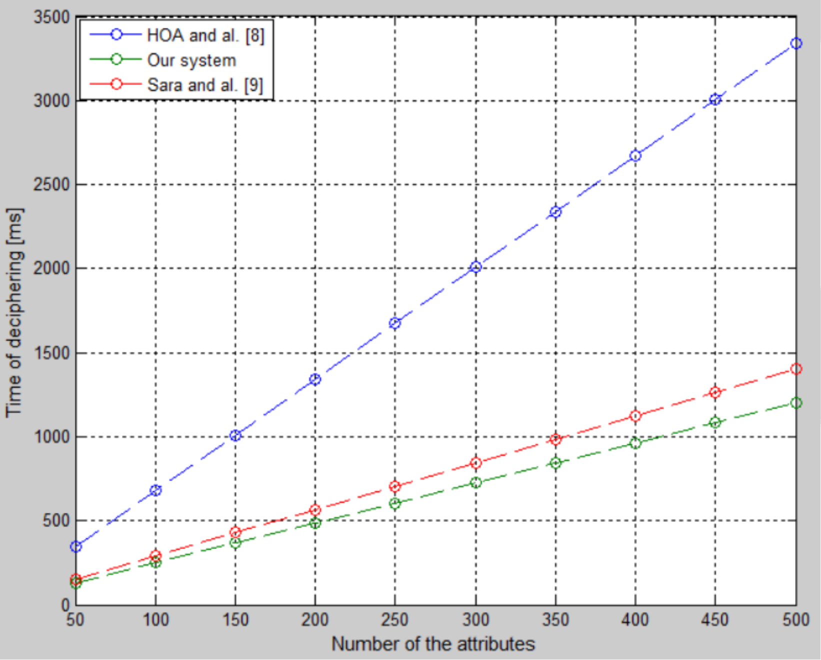


Fig- 3 : comparison of the deciphering performances of our system with other protocol.

The two figures show that the time of ciphering/deciphering is proportional to the number of the attributes. The performances of our system are though better because they spent less time in ciphering/ deciphering compared to the protocol opponents. Then, in our protocol, the combination CP-ABE with the elliptical curve has allowed us to reduce the time of ciphering and the time of deciphering.

2.4.4 Evaluation of the Blockchain strengths.

The experimental results of [10] for the Parity of Etherum customer are given in table 1

Table- 1: Performance of a private Blockchain through Ethereum with the Parity customer

Nomber of transactions (t_x)	Time for the validation T (minutes)

1000	1.74
2000	3.48
3000	5.13
4000	6.91
5000	8.71
10000	18.52

The straight model by regression can be written:

$$T = 1.9 \times 10^{-3} t_x - 3.9 \times 10^{-1} \qquad (3)$$

T : The necessary time t_x for the validation of transactions in the Blockchain

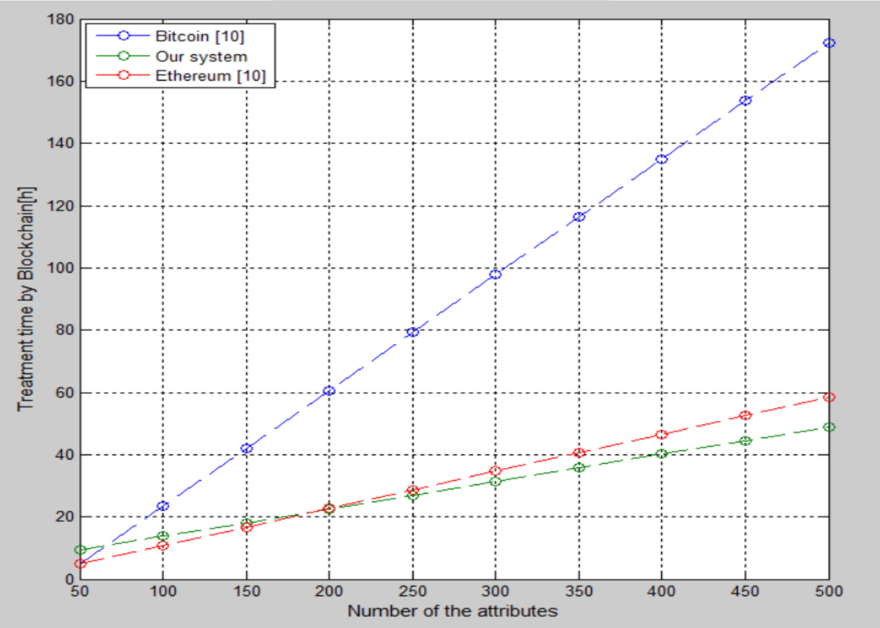


Fig- 4 : comparison of treatment time by Blockchain according to the required transactions

The figure 4 illustrates the comparison of treatment time by Blockchain according to the number of the connected attributes.

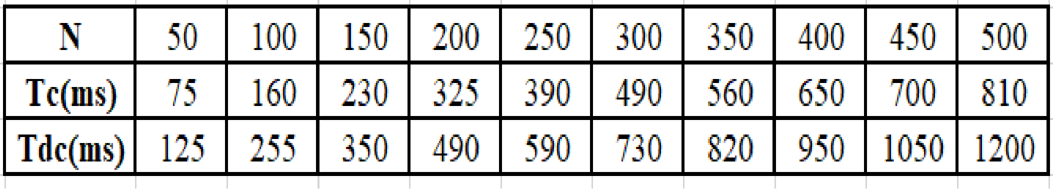
As far as we are concerned, the won results already show the superiority of Blockchain Ethereum to Bitcoin. When looking through the results in the diagram 4, if the number of connected attributes is between 50 and 200, our system shows a little difference (about 0,05%) against Blockchain Ethereum. But if the number of attributes to look over is over 200 ones, our system has the first place for Blockchain Ethereum.

3 DISCUSSION ABOUT THE RESULTS

3.1 CP-ABE combined with ECC

The simulation results of the ciphering/ deciphering algorithm are given on the following diagram.

Table- 2: Time of ciphering / deciphering according to the number of attributes.



N	50	100	150	200	250	300	350	400	450	500
Tc(ms)	75	160	230	325	390	490	560	650	700	810
Tdc(ms)	125	255	350	490	590	730	820	950	1050	1200

After doing the Analysis in Principal Component (ACP), and having calculated the Matrix of correlations U between the 3 variables N, T_C et T_{DC}, we can have :

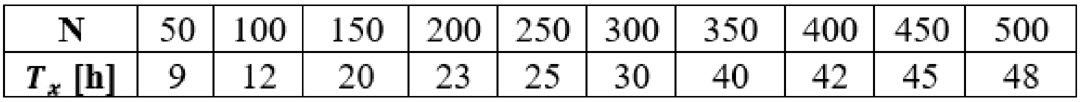
$$U = \begin{pmatrix} 1 & 0.9992 & 0.9996 \\ 0.9992 & 1 & 0.9987 \\ 0.9996 & 0.9987 & 1 \end{pmatrix}$$

According to the results of the analysis, there are high positive correlations for each of these three couples (T_C, N), (T_{DC}, N) et (T_C, T_{DC}) and the coefficients of the correlations are the following $\rho(T_C, N) = 0.9992$, $\rho(T_{DC}, N) = 0.9996$ et $(\rho(T_C, T_{DC}) = 0.9987$,

The results of the ACP justify the fiability of our ciphering and deciphering process, the value of the coefficient of determination D is 99.74%.

3.2 Our Blockchain paradigm

Table- 3: Treatment time through Blockchain according to the number of attributes



N	50	100	150	200	250	300	350	400	450	500
T_x [h]	9	12	20	23	25	30	40	42	45	48

The straight model by regression can be written:

$$T_x = 0.09N + 4.47 \quad (4)$$

The correlation is high between the two variables N et $T_{x..}$ The blockchain fiability is measured by the coefficient D which is 98,08 %. The analysis of the least squares justifies the stability of our Blockchain paradigm.

4 CONCLUSION AND PERSPECTIVES

In this article, we have proposed a ciphering by attributes CP-ABE combined with elliptical curve according to ECDH protocol to strengthen the security of the Big Data. The Blockchain paradigm allows us to organize the exchanges of Data with our dealt architecture, assuring a security of data through ciphering, and making participate the network nodes for the creation of new blockchains.

As perspective and future work, we can exploit the ant colony optimization algorithm to minimize the treatment time through the BLOCKCHAIN of our system.

REFERENCES

[1] Dissanayake, Anusha, (2021). “Big Data Security Challenges and Prevention Mechanisms in Business. International Journal of Scientific Research & Engineering Trends”. Volume 7. Issue 1. issn: 2395-566X.

[2] Mills, Lucas, Irakliotis, Ruppa, Carlson et Perlowitz. Demystifying Big Data: A Practical Guide to Transforming the Business of Government. Washington: TechAmerica Foundation, 2012.

[3] W. Diffie and M. E. Hellman. “New directions in cryptography”. IEEE Transactions on Information Theory (IT), Vol. 22, No. 6, pp.644–654, November 1976.

[4] Andreas Enge. Elliptic curves and their applications to cryptography: an introduction. Springer Science & Business Media, 2012.

[5] Simonin, X. (2016). Pour comprendre la technologie blockchain. Récupéré sur revue-banque.fr:http://www.revue-banque.fr/management-fonctions-supports/article/pour-comprendretechnologie-blockchain

[6] Cryptographie Avancée Courbes elliptiques, Cécile GONÇALVES, Décembre 2015

[7] John Bethencourt, Amit Sahai et Brent Waters: Ciphertext-Policy Attribute Based Encryption. In Proceedings of the 2007 IEEE Symposium on Security and Privacy, SP ’07, pages 321–334, Washington, DC, USA, 2007. IEEE Computer Society. ISBN 978-0-7695-2848-9.

[8] J. HAO, C. HUANG, J. NIB, H. RONG, M. XIAN, and X. SHEN. Fine-grained data access control with attribute-hiding policy for cloud-based iot. Computer Networks, 2019.

[9] M. Sara, B Lydia, Proposition d’un protocole de contrôle d’accès au big data dans le contexte de l’Internet des Objets, 2020.

[10] S. Rouhani et R. Deters: Performance analysis of ethereum transactions in private blockchain. In 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), pages 70–74, 2017.