Enhancing the Security of Data Stored in the Cloud using customized Data Visualization Patterns

Archana M[1]
Research scholar, School of CSE
REVA UNIVERSITY, Bengaluru-560064

Dr. Gururaj Murtugudde[2]
Professor, School of CSE
REVA University, Bengaluru-560064

***Abstract*—Cloud Computing is getting popularized with the invention of latest technologies like Big Data, Artificial Intelligence, Data Science etc. The biggest challenge faced by researchers is efficient ways of accessing the data and acquiring the required results. The efficiency of the system will help the researchers to go one step further in the field of cloud computing. Alongside of storing the data in an optimal way, one biggest challenge faced by the researchers is security. How best security can be enhanced for this data in order to protect the end system from data thefts and illegal attacks. In this paper the proposed research concentrates on customized data visualization techniques that have been developed in order to store the data and also enhance the data security. These visualization patterns are dynamic in nature and can be further extended based on the need and level of the security required by the application. The proposed research in this paper will help researchers to implement the data visualization techniques with enhanced security in the real time data stored in the cloud from unauthorized access and various attacks like Malware etc. and these data patterns are dynamic in nature which will be selected based on the number of fragments need to be stored pertaining to particular cluster or region. The patterns will be selected based on two factors basically, one is the number of fragments and another important factor is how many nodes are available in the pattern. This proposed research will give an additional strength to the Cloud Computing Platforms like AWS and Google Cloud where the customers can feel that their data is in safe hands. Today when we are living in the data world, the need of this system is very much essential as it enhances the security of the data.**

***Keywords*—Artificial intelligence; big data; cloud computing; data science; data visualization**

I. Introduction

Data is the today's heart of IT companies and as data is growing day by day the major issue faced is what are the best ways to store data in order to later access it easily. The data has grown from KB's to MB's and MB's to GB, today the data produced is in Zeta and Peta Bytes from each individual source. There are two major issues faced by the cloud service providers when data is growing day by day. [15] The first one is how to store this data and the second one is what the possible ways to secure the data are. On top of it how fast the data can be accessed when needed. In olden days as the size of the data was not so huge and also the data which was getting produced was not so drastic, the industries were using servers and later moved to data centers etc., The problem again here is how fast the data can be accessed and how many data centers should be established when data keeps on getting increased. So instead of working on physical resources, the concept of cloud computing has been introduced. In this cloud computing there is no limit on how much data is going to be stored and it's easy to access the data as data is stored virtually on the cloud.[3] The data which is getting generated might be structured or unstructured or semi structured. It might be string data or numerical or image data. And most of the cloud providers prefer third party to protect the data. The third party protection though it is preferred but cannot work all the times as it is difficult to trust the third party vendors. On top of it the data has to be protected again from these third party service providers. By considering all of the above scenarios and after extensive research, it is found that there is a need for a system which will help the cloud service providers easily to Store, Secure and Access the data easily from cloud [4].

The proposed research methodology utilizes the various data visualization techniques that are implemented in such a way that the data can be stored easily for providing highest level of security and also for the ease of access. The entire data will be stored in the cloud after proper conversion of quasi and unstructured data to structured data [5].

The final structured data will be stored in the form of fragments and these fragments will be stored in the form of customized patterns and preexisting patterns for providing high level security for the data [3]. In the proposed system if there is a need for third party, same can be involved without any hesitation. Though the third party is involved in controlling the cloud but they don't have any access to final data that is stored for future requirements, so the proposed system will work effectively to store and process the data when there is a need. [7].

II. Literature Survey

1) This Paper is referred to understand the data encryption techniques using code book.

2) This paper gives a clear insight on the how data visualization techniques can be implemented in light weight web pages and also scientific applications as a micro service.

3) The main take away from this paper is how data visualization techniques will be used for rain fall predictions and also understand what are the various default visualization images that can be used for this. The default visualization images are not up to the mark due to various reasons.

4) In this paper the main discussion is on how the sequential patterns will be recorded. And the sequential patterns of the objects will be stored for future data visualization patterns. It includes various objects which are static and dynamic in nature.

5) This paper gives an idea on how the data can be fragmented for easy storage. And also, how to encrypt this fragment to store in the cloud for security. And also gives an insight on how easily these fragments can be accessed.

6) In this paper the main take away is how d3 visualization can be used while decoding and storing the data. And once the data is stored how it can be extracted for future use. They built a search engine where user can give visual based inputs which will be stored in the form graphs, charts and data will be stored in that, in encoding format.

7) In this paper the main idea is how to generate evolutionary patterns using original data. These evolutionary patters are like data visualizations which will help in future for the urban planning.

8) This paper gives a clear idea on Data Visualization. Starting from what is data visualization to how these techniques can be useful in real time applications. What are the different ways this data visualization can be used like for predicting, analyzing, result, reports etc..

9) This paper is used to understand data encryption techniques using Siribhoovalaya and how data fragments can be encoded and decoded using such techniques.

*10)*This paper researches how the data stored in cloud can be utilized further for data visualization. In this paper the data taken is traffic data which can be used to generate data visualization patterns. After researching on this paper it is also clearly understood that most of the researches give much preference for data visualization while working on real-time data.

*11)*The main take away from this paper is how to work on behavioral lines which are getting generated in data visualization. These behavioural lines really help to understand the behaviors of objects or phenomena which is happening in real time. And also give insight on how to access cloud data for the same.

The literature survey on various papers gave clearly an idea to implement proposed research by avoiding few drawbacks faced by them.

III. Implementation

In order to provide high level security for the data, the data has been encrypted using the concept of Siribhoovalaya. Once the encryption process has been done the first level of security is provided for the data [5]. The next level will be how and best the data can be stored and further accessed when it is required. To do this the concept of data visualization is used in which the data will be stored in various patterns [8]. And the patterns or either the fixed or customized patterns.

Algorithm to Read and Store the Encrypted Data in Patterns:

```
1. Start
2. Read the Encrypted Data
3. Load the Encrypted Data
4. Selection of Pattern based on Selection Function
5. Store the data as per the pattern technique.
6. Verify the final pattern.
7. Stop
```

In order to retrieve the data when needed just choose the pattern saved and get back the data.

Pseudo Code to Read and Store the Encrypted Data in Patterns:

```
Procedure RS_Encrypt()
Input: Encrypted Data
Output : Pattern Based on Encrypted Data
While not end of this document
  Data[]=datastore.getinstance()
if (Data[]!=null)
        for Data[] do
count =count+1
        else
                exit()
selection (count)
Nodes=count
Select pattern based on count value
For each node do
        Store(data[])
End procedure
```

Algorithm to extract the data from the Pattern:

```
1. Start
2. Select the solution pattern from which data needs to be extracted
3. Verify once again the pattern with Select Function.
4. Retrieve the data which is stored.
5. Cross check the Encrypted data.
6. Stop
```

Pseudo Code to extract the data from the Pattern:

```
Procedure Read(data[])
Select the pattern
        For each node
                read data at every node
                 store(data[])
End Procedure
```

After these steps the data decryption algorithm will be applied in order to retrieve the original data. The main thing to observe here is that the data is not going to be stored as it is, instead, it will be stored in the form of fragments and each fragment will be assigned with a particular node in the data visualization pattern and in particular order so that it will be very easy to retrieve back the data when it is needed [9].

IV. Data Visualization Patterns

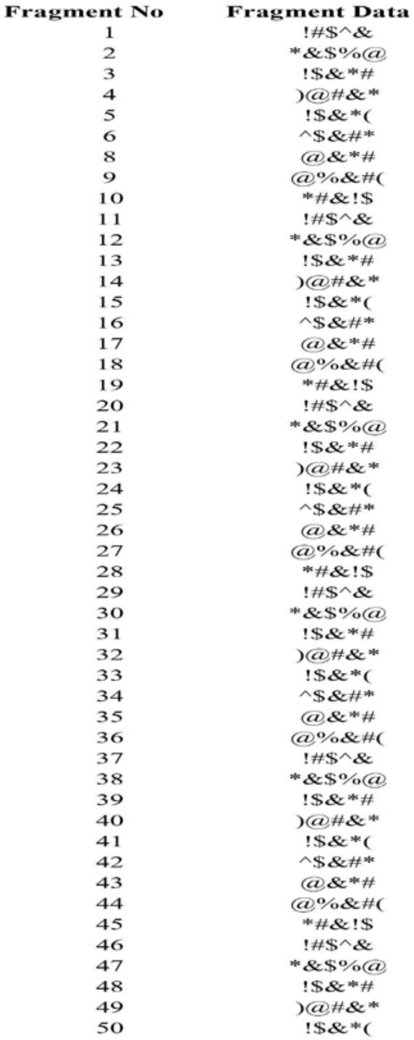
The data visualization patterns will really help in storing the data securely. And these patterns can be created and stored, which can be used basis the number of fragments and level of the security required and based on the Select Function as discussed later. Let us consider an example if the data is public that is should be shown to every customer in the organization, then normal data visualization patterns [6] will be chosen. If the data is a secured one which needs to be hid from all except the administrators, then it can use high level data visualization pattern in order to provide additional security [8].

Low Level Data visualization patterns and High-Level Data Visualization patterns will be decided basis the number of nodes available in the patterns and also it depends on the number of nodes available in the particular pattern to store the data [10].

The sample data considered here is already encrypted data using the code book algorithm pertaining to one particular category [1].

The Fragments might be a single letter or word, depending on the application it will be selected. Once the fragments are generated, in the next step data is encrypted using code book. The encrypted fragments are considered one by one to store it in the visualization pattern [12]. And for every fragment an individual ID will be generated to access it back easily when required. So the existing technology problems like Storing and Processing of cloud data will be solved in this system [1]. Also the performance of the system is tested with various factors like size of the fragment, number of fragments and Indexing. [6].

There are total 50 fragments which are encrypted. Each fragment has the encrypted data as per code book algorithm. By considering the number of fragments the following patterns are generated as a reference,



Fragment No	Fragment Data
1	!#$^&
2	*&$%@
3	!$&*#
4	)@#&*
5	!$&*(
6	^$&#*
8	@&*#
9	@%&#(
10	*#&!$
11	!#$^&
12	*&$%@
13	!$&*#
14	)@#&*
15	!$&*(
16	^$&#*
17	@&*#
18	@%&#(
19	*#&!$
20	!#$^&
21	*&$%@
22	!$&*#
23	)@#&*
24	!$&*(
25	^$&#*
26	@&*#
27	@%&#(
28	*#&!$
29	!#$^&
30	*&$%@
31	!$&*#
32	)@#&*
33	!$&*(
34	^$&#*
35	@&*#
36	@%&#(
37	!#$^&
38	*&$%@
39	!$&*#
40	)@#&*
41	!$&*(
42	^$&#*
43	@&*#
44	@%&#(
45	*#&!$
46	!#$^&
47	*&$%@
48	!$&*#
49	)@#&*
50	!$&*(

Fig. 1. Sample Encrypted Fragment Data for Data Visualization.

Based on the above encrypted fragments (Fig. 1), the below various visualization patterns (Fig. 2 to Fig. 8) are generated for the reference.

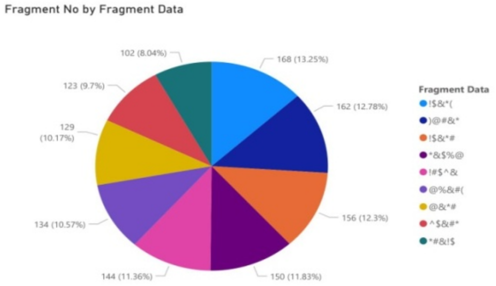


Fig. 2. Sample Data Visualization Pattern_Circular Structure.

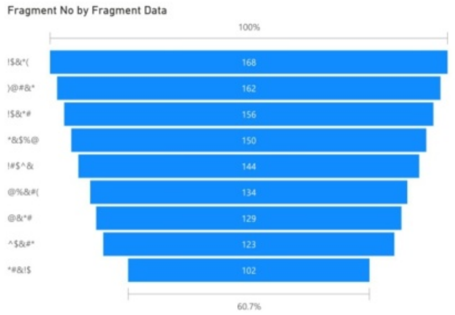


Fig. 3. Sample Data Visulaization Pattern_Sequential Steps Architecture.

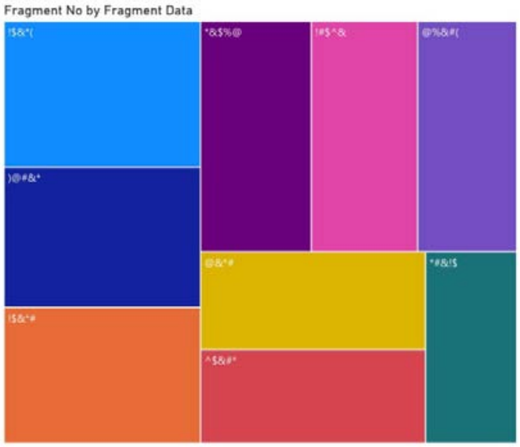


Fig. 4. Sample Data Visualization Pattern_Block Structure.

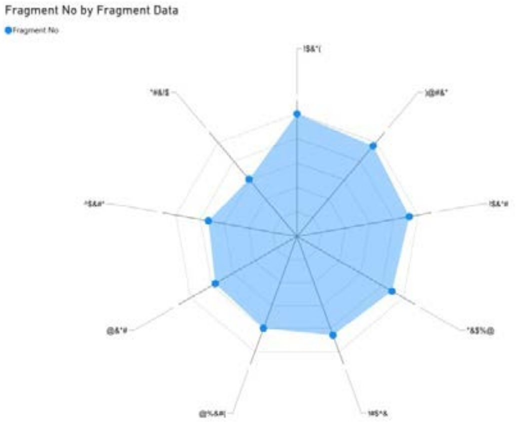


Fig. 5. Sample Data Visualization Pattern _Spiral Structure.

Fig. 2 to Fig. 8 are the sample visuals generated based on the encrypted fragment data based on the above sample data.

The above proposed method is an optimized way of storing the data into the cloud and also provides the enhanced security for the fragments in the cloud. [1].

The Encrypted Data will be stored into the cloud based on the size using either available visualization patterns or customized patterns. This data will be stored into the particular pattern based on the randomization algorithm [11].

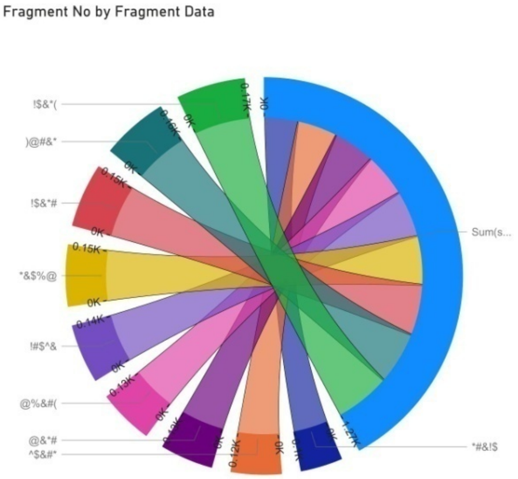


Fig. 6. Sample Data Visualization Pattern _Circular Architecture with Divisions.

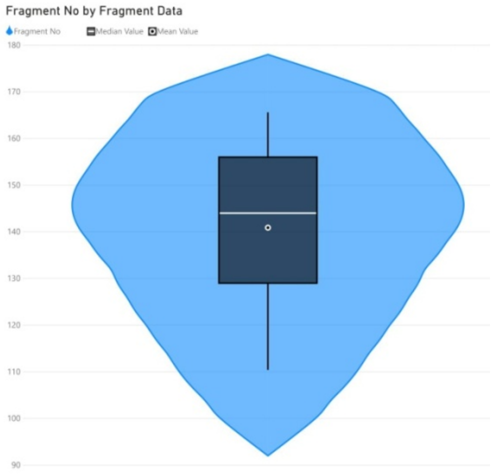


Fig. 7. Data Visualization Pattern_Leaf Architecture.

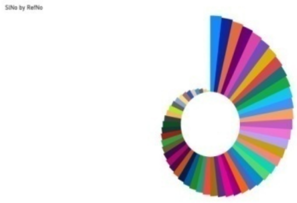


Fig. 8. Data Visualization Pattern_Snail Architecture.

Optimization

The optimization in this system is achieved by choosing the exact pattern which will store the data at various nodes depending on the size of the data. Here it is evident that the space will not be wasted unnecessarily by choosing the pattern of 100 nodes to store only three fragments data. And the level security plays very vital role while doing the optimization process.

V. Pattern Selection Algorithm for Optimization

The pattern selection algorithm in this research is used to perform the optimization process on the secured data fragments which need to be stored in the cloud for further access [13]. The pattern selection algorithm will work on the select function which is defined as f(x). Based on this select function only the corresponding pattern is selected and fragments will be stored into the corresponding location of the nodes in the patterns for further accessing when it is required [7]. The select function is defined as:

f(x)= Number of Fragments (N) / Number of nodes available in the pattern (NP)

The Number of Nodes should be always greater than Number of Fragments.

The final randomization function is defined as

$f(x) = N/NP$ where $NP > N$

This Select function will be used to decide the optimization factor while storing the data into the cloud compared to the existing methods.

VI. Customized Pattern Generation

In the previous step the procedure of selecting the existing pattern has been discussed with respect to storing the data into the cloud. While doing the above process one more limitation the system faces after quite some time is that when all the available patterns are completed then the system is going to repeat same patterns or any new patterns are going to get generated. The proposed method will generate the customized data patterns and same will be stored in the system for further use. And also there is no harm in repeating the same patterns also as the data storing will be changed though the pattern is same.

In order to generate and store customized patterns various factors will be considered. The first and foremost factor is number of fragments which need to be stored in the pattern. Then comes the level of complexity or security. The pattern generation algorithm will consider above parameters majorly while generating the pattern.

Pattern Generation Algorithm

- Start.
- The number of fragments needs to be stored (n) (the value of n starts from 1,2,3,4,5…)
- Identify the Level of Complexity (L-Low Complexity, M-Medium Complexity, H- High Complexity)
    - The complexity level can be fixed based on the level of security needs to be provided for example, L=3, M=5, H=7.
- Calculate the value of f(x), where f(x) = n*Level of Complexity
- Generate the pattern based on the value of f(x)
- Store the pattern into the system
- Stop

Once the new pattern stored into the system, again the pattern selection algorithm procedure applied in order to select the pattern based on number of fragments need to be stored.

Consider an example n=1 and L=3 then the value of f(x) = n*L = 1*3=3, means total 3 arcs need to be generated with 3 node places, out of which one is original node and remaining two will be dummy nodes.

So the default pattern will be generated and it can be customized as per the requirements. While generating the customized patterns based on the value of f(x), elements can be selected by the users like either they can use lines, boxes, QR-Codes, Circles, Some Special Images etc.,

Are the elements limited while generating pattern? Not really the user can choose various types of elements as it will be generated before storing the data, so that the system will be trained with the sample customized patterns as shown in Fig. 9, Fig. 10 and Fig. 11.

As shown above the customized data patterns can be generated by considering the f(x) value thereby storing the data further using pattern selection and randomization algorithm.

Fig. 9. Multi QR Code Based Pattern.

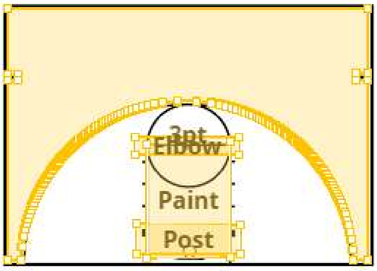


Fig. 10. Basket Ball Half Court Pattern.

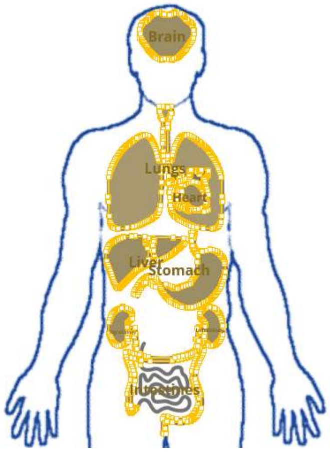


Fig. 11. Human Body – Pattern and Data can be Stored in various Parts.

VII. Comparative Study

As most of the researchers working on cloud based data access and security, most of the researchers are using graph coloring process after data fragmentation process [2]. The main drawback of this algorithms is that the complexity will increase with the increase of degree of nodes (Number of Nodes) due to which the process of storing the data will be a hectic process and accessing also not so easy which will lead to NP-Hard problem [14]. To overcome this in the proposed method the pattern selection algorithm is defined in such a way that the complexity remains same though the degree of nodes will get increased continuously. Whatever might be the degree of nodes based on select function defined in Pattern selection algorithm, the pattern will be selected but the procedure to store and retrieve the data will remain the same. That means the select function can be used with various degree of nodes represented as Δ, where the value of Δvaries from 1 to n. After the process and various experiments performed are based on Δ values that the Select function derives as.

$$f(x) = \sum_{\Delta=1}^{n} N/NP \text{where } NP>N$$

The proposed method will never lead to the NP-Hard problem as the procedure followed to generate patterns is independent to the Degree of Nodes [14].

The proposed method is very much optimized compared to the existing algorithms in order to store and access the data from cloud. Each and every pattern when storing in the cloud will be stored with a key reference for better access [5]. For example if the fragments are related to a company called A the key used to store these company fragments are represented as A_Frag_1.

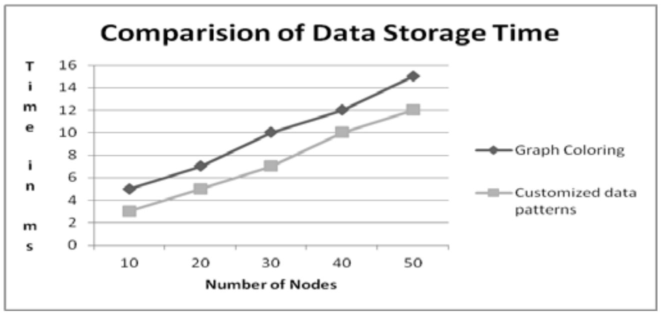


Fig. 12. Comparison of Data Storage Time.

This graph (Fig. 12) depicts the time taken for storing the data onto the cloud based in the form of number of nodes. Ex: to store 10 nodes in Customized Data Pattern technique it will take 3ms whereas to do the same in existing graph coloring algorithm it will take 5ms.

This graph (Fig. 13) represents the time taken to retrieve back the nodes from cloud. Ex: in existing graph coloring algorithm it took approximately 12-13ms to retrieve 10 nodes whereas in the proposed system it took 6 to 7 ms to retrieve 10 nodes.

Fig. 12 and Fig. 13 clearly show that the results of the proposed system are satisfactory compared to existing methods.

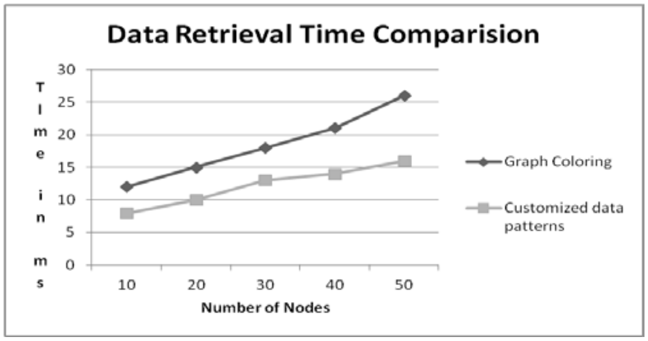


Fig. 13. Data Retrieval Time Comparison.

VIII. Enhanced Security

Coming to security as the hackers try to hack the data, the Encrypted data position is not known to the hackers as the size of the pattern i.e. number of node places will be decided by the service provider so that the select function value will be very difficult to identify. If suppose the hacker identified the random function again, he needs to work on decrypting the data as the encryption algorithm used in this paper is very secured which is called as Code Book based algorithm [1]. The proposed system compared to existing methods is optimized and enhances the security for fragments which are getting stored in the cloud. Fig. 14 gives the encryption time for different algorithms.

In the Table I the proposed Code Book (ref. Table II) algorithm is compared with the existing algorithms with respect to encryption time. And it is found that the code book algorithm will give very good results compared to existing algorithms with respect to encryption.

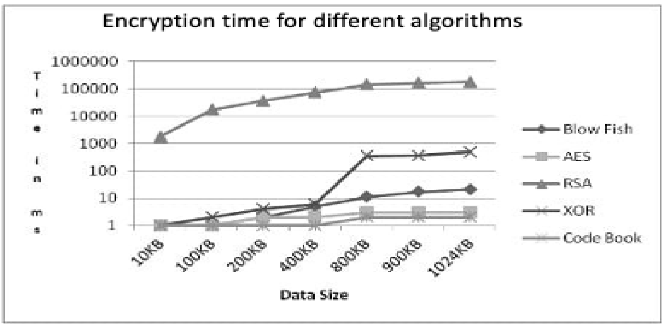


Fig. 14. Encryption Time for different Algorithms.

TABLE I. Average Time Taken for Encryption of Data Fragments (Based on Size)

Time in (ms)	10KB	100KB	200KB	400KB	800KB	900KB	1024KB
Blow Fish	1	1	2	5	11	17	21
AES	1	1	2	2	3	3	3
RSA	1743	17177	36814	73216	143206	162905	181246
XOR	1	2	4	6	356	358	489
Code Book	1	1	1	1	2	2	2

TABLE II. SAMPLE CODE BOOK

S. No	Letter/Number	Symbol
1.	A	↓
2.	a	€
3.	B	¥
4.	b	α
5.	1	ɰ̰
6.	2	ϥ
8.	3	э
9.	,	ϒ
10.	.	✓

IX. REAL TIME TESTING OF THE PROPOSED RESEARCH

In order to test the encryption technique used and data visualization techniques, the data has been shared with the certified ethical hackers to test it. The ethical hackers tried for almost 10 to 15 days and came back and said it is very difficult to crack the final data, so it is evident that the proposed research provides best security for the data which is getting stored in the cloud. The proposed method will provide multi-level security for the data which is getting stored on the cloud. One method is in the form of Encryption and another way using data visualization [6].

X. CONCLUSION

It is evident that from the proposed research whatever data is getting stored on the cloud is very well protected from all the attacks which are expected while working in real time. And also the proposed solution will give ease of access to the data which is stored on the cloud. The patterns are selected basis the pattern selection algorithm based on the number of fragments and also the number of data node representations available in the particular pattern. So it is very clear that the data storing and processing will be an easy process with the proposed procedure.

XI. FUTURE ENHANCEMENT

At present the proposed solution is working very well for the data which is stored as either individual data set or as a fragment without any limitation. In future, same system can be used to store the images for providing security. And also the data visualization patterns can be drawn in much more complex way in order to increase the level of security.

REFERENCES

[1] M. Archana, P. M. Mallikarjuna Shastry, “Hierarchical encryption algorithm to secure data fragments in cloud computing,” in International Journal of Future Generation Communication and Networking, vol. 14, no. 1, 2021.

[2] M. Raji, A. Hota, T. Hobson, J. Huang, “Scientific Visualization as a Microservice”, in IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 4, pp. 1760-1774, 2020. doi: 10.1109/TVCG.2018.2879672.

[3] Y. K. Joshi, U. Chawla, S. Shukla, “Rainfall Prediction Using Data Visualisation Techniques,” 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 327-331, 2020. doi: 10.1109/Confluence47617.2020.9057928.

[4] T. Wiktorski, A. Królak, K. Rosińska, P. Strumillo, J. C. Lin, “Visualization of Generic Utility of Sequential Patterns”, in IEEE Access, vol. 8, pp. 78004-78014, 2020. doi: 10.1109/ACCESS.2020.2989165.

[5] M. Archana, P. M. Mallikarjun Shastry, “A Method for Text Data Fragmentation to Provide Security in Cloud Computing,” International Journal of Engineering and Advanced Technology (IJEAT) ISSN, vol. 9, no. 2, pp. 2249 – 8958, 2019.

[6] E. Hoque, M. Agrawala, “Searching the Visual Style and Structure of D3 Visualizations”, in IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 1, pp. 1236-1245, 2020, doi: 10.1109/TVCG.2019.2934431.

[7] X. Shi, F. Lv, D. Seng, B. Xing, J. Chen, “Exploring the Evolutionary Patterns of Urban Activity Areas Based on Origin-Destination Data”, in IEEE Access, vol. 7, pp. 20416-20431, 2019. doi: 10.1109/ACCESS.2019.2897070.

[8] M. Islam, S. Jin, “An Overview of Data Visualization”, 2019 International Conference on Information Science and Communications Technologies (ICISCT), pp. 1-7, 2019. doi: 10.1109/ICISCT47635.2019.9012031.

[9] T. M. Aruna, M. S. Satyanarayana, G. N. Divyaraj, “A Unique Work of Out of Sight Epigraphy Creation for Data Security”, Journal of Advanced Research in Dynamical and Control Systems, vol. 11, no. 7, 2019.

[10] Jia Chaolong, Wang Hanning, Wei Lili, “Research on Visualization of Multi-Dimensional Real-Time Traffic Data Stream Based on Cloud Computing”, Procedia Engineering, vol. 137, pp. 709-718, 2016. ISSN 1877-7058. doi:10.1016/j.proeng.2016.01.308.

[11] C. Muelder, B. Zhu, W. Chen, H. Zhang, K. Ma, “Visual Analysis of Cloud Computing Performance Using Behavioral Lines”, in IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 6, pp. 1694-1704, 2016, doi: 10.1109/TVCG.2016.2534558.

[12] L. Battle, P. Duan, Z. Miranda, D. Mukusheva, R. Chang, M. Stone-braker, “Beagle: Automated extraction and interpretation of visualizations from the Web”, Proc. of SIGCHI, vol. 594, no. 8, pp. 1-594, 2018.

[13] E. Glassman, T. Zhang, B. Hartmann, M. Kim, “Visualizing API usage examples at scale,” Proc. of SIGCHI, vol. 580, no. 12, pp. 1-580, 2018.

[14] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, et al., “What makes a visualization memorable?”, IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2306-2315, 2013.

[15] "www.datapine.com [Online].