Machine Learning Based Efficient and Secure Storage Mechanism in Cloud Computing

Kanav Sadawarti, Satish Saini

Abstract**: *The cloud is an online platform that offers services for end-users by ensuring the Quality of services (QoS) of the data. Since, the user's access data through the internet, therefore problem like Security and confidentiality of cloud data appears. To resolve this problem, encryption mechanism named as Rivest–Shamir–Adleman (RSA) with Triple Data Encryption Standard (DES) approach is used in hybridization. This paper mainly focused on two issues, such as Security and Storage of data. The Security of cloud data is resolved using the encryption approach, whereas, the data storage is performed using Modified Best Fit Decreasing (MBFD) with Whale Optimization algorithm (WOA)&Artificial Neural Network (ANN) approach. The neural network with the whale as an optimization approach model makes sure the high confidentiality of cloud data storage in a managed way. From the experiment, it is analyzed that the proposed cloud system performs better in terms of energy consumption, delay, and Service Level Agreement (SLA) violation.***

Keywords: Cloud storage, Security, encryption, Rivest–Shamir–Adleman, Triple Data Encryption Standard, Modified Best Fit Decreasing, Whale Optimization algorithm, Artificial neural network.

I. INTRODUCTION

The next upcoming phase in the evolution of the internet is Cloud Computing that offers easy access to data from the infrastructure to its use as well as provided services whenever the user needed [1]. Cloud is comprised of mainly hardware, storage unit, networks along with its interfacing devices and services that enable users to access the required infrastructure, used power as per the requirement, applications, and services, regardless of location [2]. Cloud computing is a technique, which typically involves the transfer, Storage, and processing of data in the 'providers' infrastructure that are not in corporate within the user's management policy. To store data in the cloud means that saving data of users on the remote database, and the users can access data through an internet connection by making a connection to the remote database and the desired user's computer or laptop [3].The cloud computing infrastructure mainly consists of two components, such as:

- Cloud service provider (CSP): It is a management authority, which is used to control the cloud storage space with computational power.
- Cloud User: It is a person, which comprises a large number of data that is to be stored on the cloud and can be accessed later. User can be an individual or a group in the form of an organization [4].

Mainly there are three cloud usages models, such as Private Cloud, Public Cloud, and Hybrid Cloud. The cloud technology is the frequently adaptable technique due to its flexible nature of data handling. Still, it faces some challenges towards data security and hence requires an authenticated system while communicating with the users [5]. To provide Security to the stored cloud data, encryption techniques named as Rivest–Shamir–Adleman (RSA) with Triple Data Encryption Standard (DES) approach has been used. The encrypted data is stored into the database, which is further allocated to the user depending upon their priority level obtained using MBFD approach. At last, to enhance the Security of the designed cloud model, cross-validation has been performed using whale (optimization approach) with Artificial Neural Network (ANN) as a classification approach.

II. RELATED WORK

In this section, the recent research presented by several researchers in cloud security is given. In this research, we mainly discussed two aspects, such as security issues in cloud computing and the key active methods used to cloud Storage. Initially, the key security threats found in cloud data storage are discussed. On the other side, points to the limitations of current data retention techniques, which proves the need for our proposed research is considered.

Lee et al. (2018) implemented Heroku as a platform of cloud, after that introduced AES to make data secure in Heroku. Advanced Encryption Standard (AES) has been seen as the most efficient symmetric technique. Authors have focused on AES- 128 bit to make data encryption. This work has been done in four defined steps named as Substitute Bytes, shift-Rows, Mixing Columns, Add-Round-Key performed one after another. While performing research work to access data in Heroku stored in the database, a username and unique password are required for authentication [6]. Ebrahim et al. (2018) presented a security model to protect data of cloud from unauthorized access through combining algorithm i.e., steganography and cryptography. The proposed model to preserve privacy is included three phases (i) hash data by using secure hash algorithm (SHA), (ii) encryption of data by utilizing AES algorithm and encrypt hash code obtained a new key by applying public-key cryptography and at the end (iii) least significant bit (LSB) steganography has used to combined all information after encryption in any kind of images [7].

Revised Manuscript Received on March 30, 2020.
* Correspondence Author

Kanav Sadawarti*, Scholar, Computer Science & Engineering, RMIT, Mandi, Gobindgarh, India. Email: kanavsadawarti.sm@yahoo.com

Dr. Satish Saini, Electronics communication & Engineering, RMIT, Mandi, Gobindgarh, India. Email: satishsainiece@gmail.com

Timothy, D. P., & Santra, A. K. (2017) developed a novel security scheme by combining Blowfish symmetric and RSA asymmetric cryptographic algorithm and also a digital signature on the data transmission phase to enhance data security in the cloud. To accomplish the secure transmission for confidential data is the main aim of this proposed scheme. The combined structure of symmetric and asymmetric algorithms produces improved efficiency, and Security is increased to transmit data online by using the SHA-2 algorithm [8].Amalarethinam, I. G., & Leena, H. M. (2017) proposed enhanced RSA (ERSA) algorithm by using two extra prime numbers in Standard RSA algorithm. By the utilization of prime numbers instead of random numbers in the proposed work enhanced speed of encryption and decryption. This speed is constant in ERSA through dividing the file among various blocks. Apart from this benefit, the implementation of ERSA algorithm reduces the computation complexity and improved security [9]. Sadkhan, S. B., & Abdulraheem, F. H. (2017) evaluated the RSA cryptosystem through adaptive neural fuzzy inference system (ANFIS) in MATLAB. This ANFIS scheme is combination of Artificial Neural network (ANN) and fuzzy logic algorithm. To analyze the performance of this proposed algorithm, five different parameters has been utilized namely; key-length, time-complexity, entropy of cipher-text, size of plain-text, and the entropy of private key. The obtained time complexity after evaluation is 0.1357 in sec is represent the times to compute in RSA algorithm [10].Ardy et al. (2017) introduced a combined result of three algorithm named as; Rivest-Shamir-Adleman (RSA), vigenere cipher and message digest-5 (MD-5). To observe the reliability of this algorithm, this method tested with multiple attacks named as blurring, salt, and pepper, Gaussian filters etc. As per the attacking result, the smallest changes have been occurred in case of blurring attack having very good peak (signal to noise ratio) PSNR value is 86.7532 dB [11]. Arora et al. (2017) proposed a hybrid cryptographic system which gives combining advantages of both symmetric and asymmetric type of encryption by which produce secure cloud environment. This work has focused on creating a secure cloud ecosystem in which utilizes a multifactor authentication scheme, including more than one level of encryption and hashing. This presented work along with comparative algorithm has been simulated on CloudSim simulator [12].Ubale et al. (2017) introduced a role-based access control (RBAC) model, including AES and RSA algorithm, which has been provided to accomplish enhanced Security. The asymmetric key algorithm is utilized only for encrypting the symmetric key, and it needs a very small cost of computation. In this combined approach of AES-RSA, RSA is utilized here for decryption of key used in AES and also solves the problem of key transport and provides improved performance. Here, the key-pairs produced by the receiver through an asymmetric key algorithm, and the public key is distributed among senders [13].

III. PROPOSED MODEL

The proposed security model designed for cloud data storage is described in this section. The work has been executed in two phases, namely (i) data storage using encryption approach and (ii) Security of the stored data suing whale and ANN approach. The flow of the work is shown in Fig. 1.

Initially, the data is collected from n number of users and then processed data by applying the stop word removal approach.

A. Stop word removal

It is a dictionary-based approach and is used to remove commonly used word having no meaning such as, has, have, is, am, and many more from the accepted data from the users. The algorithm for stop word removal is defined below;

Algorithm 1:Stop Words Removal

Required Input:	UD← User data according to the users
Obtained Output:	SWFD← Stop Words Free Data

```
Start
Upload List of Stop Word (SW)
For each element in UD
For each element in SW
      If UD=SW
          SWFD=UD
Else
          SWFD=‘ ’
      End – If
End – For
End – For
Return: SWFD as a list of stop words free data
End – Function
```

After applying the pre-processed approach, the data is passed to word to vector approach. In this approach, the text document is converted into a numerical form. This is possible by calculating the ASCII code of each word present in the text. On each value of ASCII code, logarithmic value is finding out by using the following formula;

$$L_v = \int_0^n \log_{10}(\log_2(x))$$

Here, L_v→ Logarithmic value
n→ number of documents
x→ ASCII value

After computing the logarithmic value of each ASCII code, the average value is determined using;

$$\sum_{i=1}^{n} \frac{L_v}{n}$$

If the computed average value is less than the threshold value, then apply the RSA approach for encryption; otherwise, use a 3 DES approach for encryption.

The encrypted data is then provided as input to the Modified best fit decreasing (MBFD)approach. This approach ordered the tasks as per their power consumption. The algorithm for MBFD is written as algorithm 2;

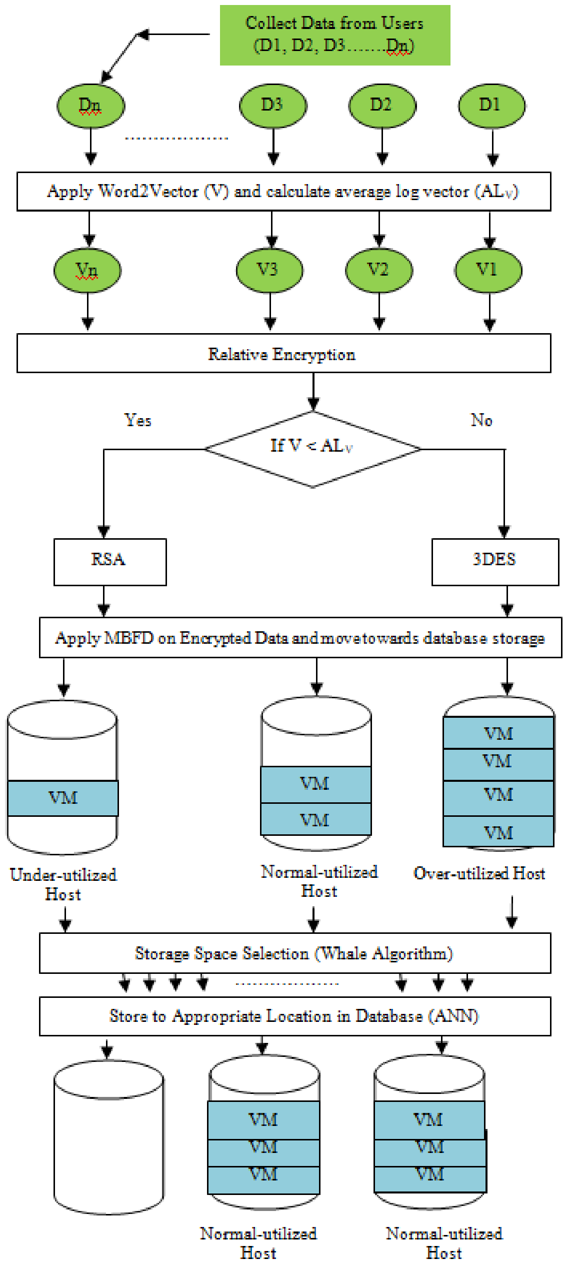

Collect Data from Users (D1, D2, D3…….Dn)
Dn
D3
D2
D1
Apply Word2Vector (V) and calculate average log vector (ALV)
Vn
V3
V2
V1
Relative Encryption
Yes
If V < ALV
No
RSA
3DES
Apply MBFD on Encrypted Data and move towards database storage
VM
VM
VM
VM
VM
VM
VM
Under-utilized Host
Normal-utilized Host
Over-utilized Host
Storage Space Selection (Whale Algorithm)
Store to Appropriate Location in Database (ANN)
VM
VM
VM
VM
VM
VM
Normal-utilized Host
Normal-utilized Host

Fig. 1. Flow of proposed work

Algorithm 2: MBFD

Required Input:	V← Word to vector data according to the users EV← Encrypted Vector
Obtained Output:	Sorted ED ← Sorted Encrypted Data

```
Start
For each element in V
{Calcuate, MinPriority ← max
    Allocated Storage ← [null] }
For each EV
{Priority, P ← EstimatedPriority (EV, V)
If P<mean(P) then
{Allocate Storage ← ED
Min(P) ← P}
  if Allocated Storage ≠ null then
  {Allocate Storage
              Sort ED according to P}
  End – If
  End – If
  End – For
  End – For
  Return: Sorted ED based on their priority
  End – Function
```

To stored data in a managed way, we have used whale as an optimization approach. Using the optimizations approach, only those VMs are migrated that are demanded by the user.

The whale is one of the newest optimization algorithms that came into existence by 2016. The algorithm works on the mechanism of the humpback nature of whales. Initially, the whales identify the position of prey and then encircle that particular location. As the location in the database is not known prior, therefore, the whale algorithm assumed the location of prey is the best optimum solution. Then the search agent updates the current position and tries to reach towards the target that is the best data space into the available dataspace. They used whale algorithm with ANN is defined below;

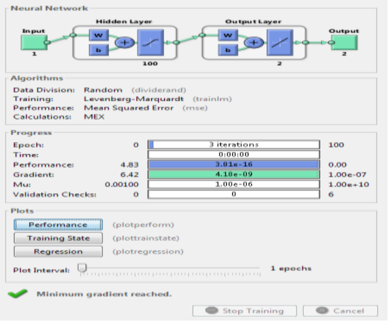
Algorithm 3: Hybridization of ANN with WOA (WOA-ANN)

Required Input:	Sorted ED-Data ← Properties of Sorted ED Cat ←Target/Category in terms of over-loaded, under-loaded and normal Storage N ← Carrier Neurons Number
Obtained Output:	Storage OL and UL ←Identified over-loaded and under-loaded Storage

```
Start
To optimized the Sorted ED-Data, Whale Optimization
    Algorithm (WOA) is used
Set up basic parameters of WOA:
        WhaleSize (W) – Based on the number of sorted
    ED
        OW – Other Whales
        Optimized Sorted ED-Data – Optimized Training
Data
        Fitness Function:
$F(f) = \begin{cases} 1\ (True); & if\ W_c < W_t = Other\ Threshold_{Properties} \\ 0\ (False); & Otherwise \end{cases}$

Where    Wc: It is properties of current ED (Current
Whale) which are in T-Data and
$W_t$: It is the threshold properties of all ED on the basis of
priority with respect to the
Calculate Length of Sorted ED-Data in terms of R
Set,Optimized Training Data, Optimized Sorted
    ED-Data = []
For i =1→Length(Optimized Sorted ED-Data)
    $W_c$ = T (i) = $Selected\ ED_{Properties}$ // Current Data
    from ED List
    $W_t$ = $Threshold_{Properties}$ // Average OW
    $Fit(f) = Fit\ Fun\ (W_c, W_t)$
    $Best_{Prop}$ = Optimized Sorted ED-Data = GOA (Fit(f),
    Sorted ED-Data, Set up of GOA)
End – For
```

```
ANN Initialization using the following parameters –
    Number of Epochs (E) // Iterations used by ANN
 – Number of Neurons (N) // Used as a carrier in ANN
– Performance: MSE, Gradient, Mutation, and Validation
–Techniques: Levenberg Marquardt
– Data Division: Random
For i = 1 →Optimized Sorted ED-Data
    If Optimized Sorted ED-Data is a subset of
    over-loadedstorage space
        G (1) = Optimized Sorted ED-Data (i)
Else if Optimized Sorted ED-Data is a subset of
    under-loaded storage space
        G (2) = Optimized Sorted ED-Data (i)
Else
    G (3) = Optimized Sorted ED-Data (i) // Normal
    storage space
End – If
End – For
Initialized the ANN using Training data and Group
Storage-Net = Newff (Optimized Sorted ED-Data, G,
    N) // Call the initialization function of neural network
Set the training parameters according to the
    requirements and train the system
Storage -Net = Train (Storage -Net, OT-Data, G)
Verification of Model:
Current SensorNode = Properties of current sensor node
Verification Result = simulate (Storage -Net, Current
    Sensor Node)
If Verification Result = True
Consider for normal storage space
Else
Storage OL and UL = Storage Space
End – If
Return: Storage OL and UL
End – Function
```



Data Division: Random (dividerand)
Training: Levenberg-Marquardt (trainlm)
Performance: Mean Squared Error (mse)
Calculations: MEX

Epoch: 0 3 iterations 100
Time: 0:00:00
Performance: 4.83 3.81e-16 0.00
Gradient: 6.42 4.18e-09 1.00e-07
Mu: 0.00100 1.00e-06 1.00e+10
Validation Checks: 0 0 6

Performance (plotperform)
Training State (plottrainstate)
Regression (plotregression)

✓ Minimum gradient reached.

Fig. 2. Trained ANN structure

WOA helps to select an appropriate area in the database based on the fitness function. The optimized user's data is then provided as input to the ANN layers, which trains the system to find out the over and underutilized database space. The trained structure of ANN is shown in Fig. 2.

IV. RESULTS AND DISCUSSION

To determine the performance of the designed security cloud model, the parameters in terms of delay, energy consumption, and SLA violation have been evaluated with MBFD and with WOA & ANN technique. The computed results are shown in Table 1.

Table -1: Energy Consumption mJ

Number of users requests	With MBFD	With WOA &ANN
10	8	6
20	10	7
30	14	11
40	17	15
50	19	16
60	22	20
70	28	25
80	33	29
90	37	33
100	44	39

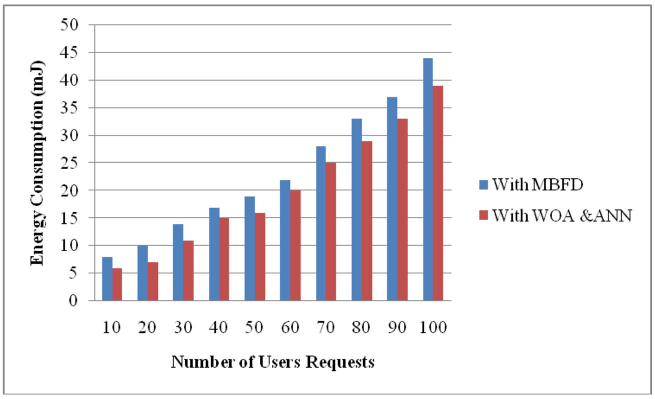


Fig. 3. Energy Consumption

Energy consumption in the cloud environment is the most relevant parameter, which must be minimized to save energy. Here, the parameter is examined with respect to the number of user's requests. The user requests vary from 10 to 100 and hence analyze the energy consumed to allocate the tasks at an appropriate space within the database. From the graph, it is very clear that as the user's requests increase, the energy consumption also increases. The average energy consumed with the MBFD approach and with WOA & ANN approach is 23.2mJ and 20.1 mJ, respectively. Thus, there is a reduction of 13.36 % in energy consumption while utilizing WOA with the ANN approach.

Table- 2: Delay (s)

Number of users requests	With MBFD	With WOA &ANN
10	0.9	.5
20	1.5	.8
30	1.9	1.1
40	2.4	1.5
50	2.8	1.7
60	2.9	2.2
70	3.4	2.4
80	3.7	2.8
90	3.9	3.4
100	4.4	3.6

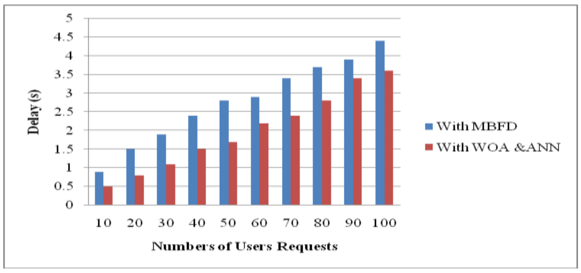


Fig. 4. Delay

To observed and examined that the uploaded data in the cloud is secure or not and also allocate appropriate storage space, the delay parameter is examined. Practically, in a cloud environment, it is seen that the data traffic is increasing day by day and will become high that influences the performance of the cloud system. The delay in cloud occurs due to many factors such as a number of user's requests, size of uploaded data, speed of the network, and so on, which results in delay or congestion during communication. Here, efforts have been made to reduce the delay by applying WOA with ANN as a machine learning approach. It has been analyzed that the average delay observed using simple MBFD and with WOA & ANN approach is 2.78 s and 2 s, respectively. Thus, the speed of storing data into the database with an accurate manner is increased by 39 % while using WOA with the ANN approach.

Table-3: SLA Violation

Number of users requests	With MBFD	With WOA &ANN
10	8	6
20	11	7
30	14	11
40	16	15
50	17	13
60	19	15
70	22	18
80	24	20
90	26	22
100	28	24

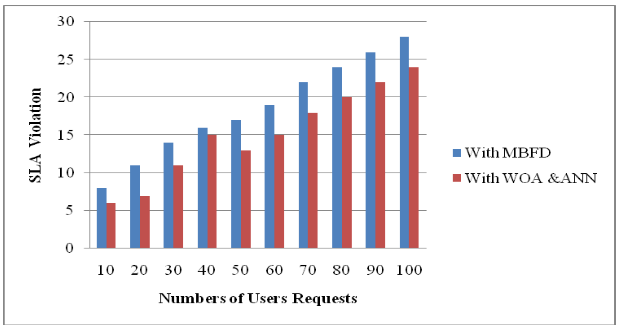


Fig. 5. SLA Violation

To make sure that the services provided by the CSP are as per the client's demand, Service Level Agreement (SLA) plays an essential role. If the services are violated by the CSP, then the providers must have to pay penalties to their users. To analyze and reduce SLA violation WOA with ANN approach is used in this research. From the graph depicted in Fig. 5, it is clear that the proposed approach provides better services to its users. The average SLA violation observed with MBFD and with WOA &ANN approach is 18.5 and 15.1, respectively. Thus, 18 .38 % reductions in the SLA violation have been examined with WOA &ANN approach compared to the simple MBFD algorithm.

V. CONCLUSION

In this article, we have presented novel security with managed storage space for a cloud computing environment. The security structure has been provided using RSA with a 3 DES encryption approach. The encrypted data is then sorted using the MBFD approach as per their energy consumption order. The storage space selection problem has been solved using the WOA approach. The appropriate selection of space in a database that is to select that space that is underutilized has been performed using ANN as a classification approach. In this way, a secure and efficient storage space cloud model has been designed. The performance has been measured based on energy consumption, delay, and SLA violation. The proposed cloud model performs well compared to a simple or traditional cloud computing system.

REFERENCES

1. S. Namasudra, “An improved attribute-based encryption technique towards the data security in cloud computing,” *Concurrency and Computation: Practice and Experience*, vol.*31, no.* 3, 2019, pp. 4364.
2. J. Li, Y. Zhang, X. Chen, & Y. Xiang, “Secure attribute-based data sharing for resource-limited users in cloud computing,” *Computers & Security*, vol. *72*, 2018, pp.1-12.
3. S. Kumar, J. Shekhar, & J. P. Singh, “Data security and encryption technique for cloud storage,” In *Cyber Security*, (2018, pp. 193-199). Springer, Singapore.
4. N. Mishra, T. K. Sharma, V. Sharma, & V. Vimal, “Secure Framework for Data Security in Cloud Computing,” In *Soft Computing: Theories and Applications* , 2018, (pp. 61-71). Springer, Singapore.
5. Goyal, & C. Kant, “An Effective Hybrid Encryption Algorithm for Ensuring Cloud Data Security,” In *Big Data Analytics* , 2018, pp. 195-210). Springer, Singapore.
6. B. H. Lee, E. K. Dewi, & M. F. Wajdi, “Data security in cloud computing using AES under HEROKU cloud,” In *2018 27th Wireless and Optical Communication Conference (WOCC)* , 2018, April, pp. 1-5). IEEE.
7. M. A. Ebrahim, I. A. M. El-Maddah and H. K. Mohamed, "Hybrid model for cloud data security using steganography," *2017 12th International Conference on Computer Engineering and Systems (ICCES)*, Cairo, 2017, pp. 135-140.
8. D. P. Timothy and A. K. Santra, "A hybrid cryptography algorithm for cloud computing security," *2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS)*, Vellore, 2017, pp. 1-5.
9. I. G. Amalarethinam and H. M. Leena, "Enhanced RSA Algorithm with Varying Key Sizes for Data Security in Cloud," *2017 World Congress on Computing and Communication Technologies (WCCCT)*, Tiruchirappalli, 2017, pp. 172-175.
10. S. B. Sadkhan and F. H. Abdulraheem, "A proposed ANFIS evaluator for RSA cryptosystem used in cloud networking," *2017 International Conference on Current Research in Computer Science and Information Technology (ICCIT)*, Slemani, 2017, pp. 48-51.
11. R. Damara Ardy, O. R. Indriani, C. A. Sari, D. R. I. M. Setiadi and E. H. Rachmawanto, "Digital image signature using triple protection cryptosystem (RSA, Vigenere, and MD5)," *2017 International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS)*, Yogyakarta, 2017, pp. 87-92.
12. A. Arora, A. Khanna, A. Rastogi and A. Agarwal, "Cloud security ecosystem for data security and privacy," *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, Noida, 2017, pp. 288-292.
13. S. A. Ubale, S. S. Apte, & J. D. Bokefode, “Developing Secure Cloud Storage System Using Access Control Models,” In *Proceedings of the International Conference on Data Engineering and Communication Technology* , 2017, pp. 141-147). Springer, Singapore.