KEYWORD BASED AUTOCLASSIFICATION OF DATA IN CLOUD COMPUTING

DAYANANDA LAL N[#1], Dr. SENTHIL KUMAR K[*2], GAUTHAM B S[#3], NARASIMHA MURTHY.M. S[#4]

[#1]Research Scholar, Department of Electronics and Communication Engineering,
Dr. MGR Educational and Research Institute University,Chennai,India.
dayanandlal@gmail.com

[*2]Professor, Department of Electronics and Communication Engineering,
Dr. MGR Educational and Research Institute University,Chennai,India.
ksenthilkumar@drmgrdu.ac.in

[#3]MTECH, 2nd Year Student, Dept. of CSE,
Acharya Institute of Technology, Bangalore, Karnataka, India.
gautham4gowda@gmail.com

[#4]Research Scholar, Dayanand Sagar College of Engineering, Bangalore, Karnataka, India.
china.mur@gmail.com

***Abstract*—There is much thinking nowadays about unsanctioned access to personal information. Different strategies have been suggested in survey process, since authors suggest augments with respect to expenditure and process evaluation instance therefore it depends on ciphering the complete information. This article expresses an introduction to cloud computational schema which differentiates information reckoning on significant properties. Most precisely the significant user information would get encrypted with secure encryption algorithm provided with better key size, where in lower significant information may not be ciphered. Apart from it introduces a new method wherein the auto classification of the data is done based on the keyword specified, if the file to be uploaded matches the specified keywords then the file is auto classified to the matched classification level , later encryption takes place and is uploaded to the cloud. This methodology helps debasing the process charges and complicatedness of information storage, since there is no compulsion for applying the similar refined encryption procedure for all user records. The outcome of introducing suggested schema results in exhibition of amelioration and effectiveness than further open frameworks. There are few of the embellishments chosen and it can be known in the following pages basically to suit the modern demands.**

Keyword-AES (Advance Encryption Standard),DES (Data Encryption Standard), TLS (Transport Layer Security),SHA (Secure Hash Algorithm)

I INTRODUCTION

User data are considered to be highest value for both individuals and for the business managements. It should not be in the position to lose the data. Hence, it results in rapid demand for safeguarding the information. At present, the solution for this is by storing the data on to the cloud and their esteems are augmenting rapidly because of reserve and self -regulating backing up of random type of information which is measured as clearly budgetedor economical, convenient to utilizeand also at ingress. This rendered a convenient way to share and synchronize the respective user data's which is in-between the users and gadgets. Businesses purpose users are afraid to store their data on cloud due to lose of control of their data which results in second thought before executing. Apart from that, current progressive attacks on the cloud reserve has lead to the worst-case concerns. To convenient the whole process, lots of solutions are defacto provided which is used to maintain the user's information more in a safeguarded manner. This paper suggests a consistent cloud schema which introduces privacy and even reliability in the corresponding data with respect to both the storage and transition of the information .It minimizes the processing time in the respective encrypted type data. It does offer improved usage of the bandwidth by segregating the data in congruence based on the level of unlockingconsidered. User always do have the concerns about storing susceptible kind data on the respective cloud service provider servers. Users are completely unaware whether the servers are considered to be secure to reserve their data because these servers may be opened to different threats, cyber-attacks.

Hence, many of respective cloud service providers [5],[6],[7],[8] has presented the different ways to manage the information a secure and safer manner. Many approaches of encryption may be considered as a better way to store the data on to the cloud system. Whereas utilizing the similar encrypting algorithm to store the user data with a constant key size is not a secure manner to store. Apart from regular data there are few data which requires higher protection. Here, we are suggesting an effective and safeguarding schema for the cloud computing reserve environment. This schema differentiates primarily between the data reckoned on their level of confidentiality and we consider distinguishing of data reckoned on the level of secrecy maintained. Also known with respect to importance, which is on the basis we perform the selection of the appropriate algorithm with also appropriate size of the key to offer for the required level of safe guarding. In this manner, we debase the price, complicatedness and decrease the time necessary to reserve the information in a secured manner.

II ANALYSING THE RESERVE AT CLOUD

Cloud storage has become the model of the evaluation in cloud where in the data is reserved on the respective servers of remote and can be accessed via Internet. The data is maintained by the reserve of cloud service providers on the corresponding servers. Minimum one single cloud server connected to Internet is required to store data. User uploads and stores the data on to the server through Internet and the server does reserves the data. If the user wishes to retrieve back the data by connecting the server of data via interface of web application type or trough the set of applications. The client can either reserve the pack of files or gain the ingress to the respective data primarily to amend on corresponding server. Cloud reserve system primarily reckons on the server in which the data is stored. The data are reserved on multiple servers primarily to ensure a sustainable service to the customers, so that at any point of time they can ingress data [9][10][11].

III BENEFITS WITH FRAMEWORK

Cloud schema has many of the benefits and does enhance the teamwork and does collaborate the efforts by permitting team members to ingress common type of data. Regardless of other benefits, user can gain from the cloud type reserve services but it does comprise of restraints. Users permission is not mandatory even without it data can be made void, rectified and even progressed further. with the help of software's provided by service providers it makes easy to setup the backup type schemes primarily on primary devices and then store all the data on the reserve servers. Similar procedure to the registrations and login, attackers may make or consider the stealing those identifications of the user from insecure server. Attacker may be get access and alsoadminister the content of the privately held data. Hence, server must or need to scrutinize itself to the respective user and it is essential to utilize the right algorithm relating to cryptography basically ciphering all types of the transmissions performed in to check for consistency which need to be affirmed.

At present there are lot of companies competing in order to provide a quality service to their clients .There are quite lot of companies which provide the services in a good quantity in their respective fields to the clients. Some of the widely used corporation product in this cloud field are Google Doc, which provide the customer the service of storing the data on the cloud corresponding data servers of the Google and provides the facilities of sharing,read access, write access etc. Mail reserve providers considering like Google, Yahoo basically allow in logging in to their email from any of the respective devices within the broadband connection. There is also an instance of the digital picture reserve offered by an Instagram, Flickr and also Picassa. Also there are many such video upload services.

Earlier days, when the cloud reserve were considered much, users were reckoned on their computers for storing data which used to act as backup if in case of any failure. However, safeguarding the information was merely based on the user since the owner was responsible for complete administration. But nowadays all the users are dependent on storing their data on to the servers via internet. Hence it may lead for both internal and external attack. An external type attacks wherein the hackers makes out the weakness in the security and does considers stealing information. The internal threats do happen when the cloud reserve offering persons utilizes the data in a precise manner. For these known reasons, encrypting reserved data are much required basically to guarantee that data which is maintained secrecy type.

Cloud reserve providers do offer their users a generally known encryption for their respective data utilizing private key which is known only to them. The common encryption offered by them may safeguard information from those exterior attacks but cannot safeguard from local interior attackers who can ingress the key from an attack of encryption key kind theft. Therefore it is essential to encrypt the data from the client system. That is it is essential to encrypt the data in the client side before transmitting and storing on to the cloud.

Symmetric key encryption is considered as an effective and efficient among other encryption algorithms. It makes use of private key to encrypt and decrypt the message. The private key is used between the receiver and sender to prevent from being hacked from the outsiders, this key is a derived key of regulation which must be expectedly at random. manner [9][11][17].

On other side, since with larger use of an internet and computing in handphone, there is an shortly erupting field called the MCC which permits the users of handphone for gaining from the corresponding cloud services. Use of the cloud reserve and computation is actually the greater plus feature for the handphone users.[18][19], Many of the authors has expressed a handphone cloud evaluation model primarily which need to utilized in the large scale networks and which as the model need to be utilized in high level networks similar to healthcare applications. In [20], a scalable kind cloudlet as the MCC model was executed and also considered to extending [21] for the massive deployment of the respective cloudlets. [22][23][24] has provided a few solutions.

IV ENHANCEMENT OF DATA SECURITY USING RSA

Somani et al tried to guesstimate cloud reserve modus operandi and security level of data in the cloud via enacting RSA algorithm along with the digital signature to encrypt respective information before being transferred. They are defined with the digital type of signature as a format of arithmetic to get a evident of digital kind message in authentication process. Enacting a proper type of digital signature combining along with the RSA kind algorithm does guarantee that the data is secure on cloud and it permits the receipt to authenticate the message for a particular sender.

V CONSIDERATIONS FOR CONVENIENCE

For an image type encryption, blowfish algorithms are considered. Address classification is reckoned on the respective keywords. Performance chart, which is known, is based on the constraints of time and key. When considering user part, an user can download the elementary and the confidential type file in a convenient manner.

VI STORING WITH SAFENESS

Storing the digital type asset in the safer manner. Zhang et al [25] has suggested Cloudsafe to embellish the availability and the even unrevealing reserved information in its respective cloud via an encryption and making to encode the respective data into several of the providers of the cloud. In order to bring to safeness, a dependable and quick type data in ingress to possible repository, Cloudsafe brings forward cloud reckoned digital asset secure kind of service which gives out important assets primarily between many cloud service providers by utilizing erasure kind of programming and cryptography. Some of the reserve providers are Dropbox, Google Drive, Microsoft SkyDrive and also the SugarSync. In accordance to Zhang et al.[25], the availability of those augments by utilizing the erasure coding primarily to disperse the related information on many cloud providers, primarily to recover those ingress of data when the related service providers fails. AES[28] is used for encryption and decryption of their respective information to manage the unrevealing of it.

VII CONSTRAINTS OF SOLUTIONS

In each of the solutions, there are certain restrictions. For an instance, as per the research in [25], a lot of management effort is required to disperse the data on to multiple servers. They also utilize the AES type encryption which are considered costly since it does result in the extra time in processing and the corresponding assets. It also requires for the third party which does restricts the administering of data. Moreover, setting the condition on third party data will affect the privacy of the data. Lin et al. [22] and Zhang et al. [25] has proposed same restraint which is preserving data on many of the servers.

VIII SUGGESTED SCHEMA

This particular section exhibits the ECCSF which does encrypts the respective data in accordance to their unrevealing scale in three different levels. Basic, confidentiality and the greater level of unrevealing. The suggested results is reckoned on the proposal of the information segregation wherein respective user becomes liable for defining the secrecy degree of the relevant data.

IX AUTO CLASSIFICATION OF THE DATA

In today’s world most the user makes use of cloud in order to store their data and retrieve data from the cloud. Here we have introduced an approach wherein we try to classify the data based on the level of the confidentiality from the user. User can classify their data as Basic level, confidential level and highly confidential level.Based on the classification from the user the data are encrypted by applying the encryption algorithm, store the data on to the cloud and decrypt the data from the cloud. In order to download the data secret key is used to decrypt the encrypted file.One advance approach introduced here is auto classification of the user data/files. In this approach data is classified automatically based on the type of the file that is to be uploaded.

When we know it is the automatic data classification, operating system self-regulating segregates a newly created file by administering the application that regulated the respective file to amend the file by considering one or more settings for usage of data in attributes to the file in advance to the application reserving the file in a respective folder. The whole process is like aself regulating detection, categorizing and also authorized data which is of personal type.AES in general does specify FIPS approved algorithm which are cryptographic and which can be used to safeguard the electronic type data. AES can be represented as the symmetric segment of cipher, which can encrypt and even decrypt the respective information. There is conversion of data to a better form representing those of cipher text. The classification occurs based on the set of keywords specified. Whenever user try to upload the data on to the cloud the framework access the file, searches the content, if the content matches with the set of predefined keywords then the auto classification of the data takes place.

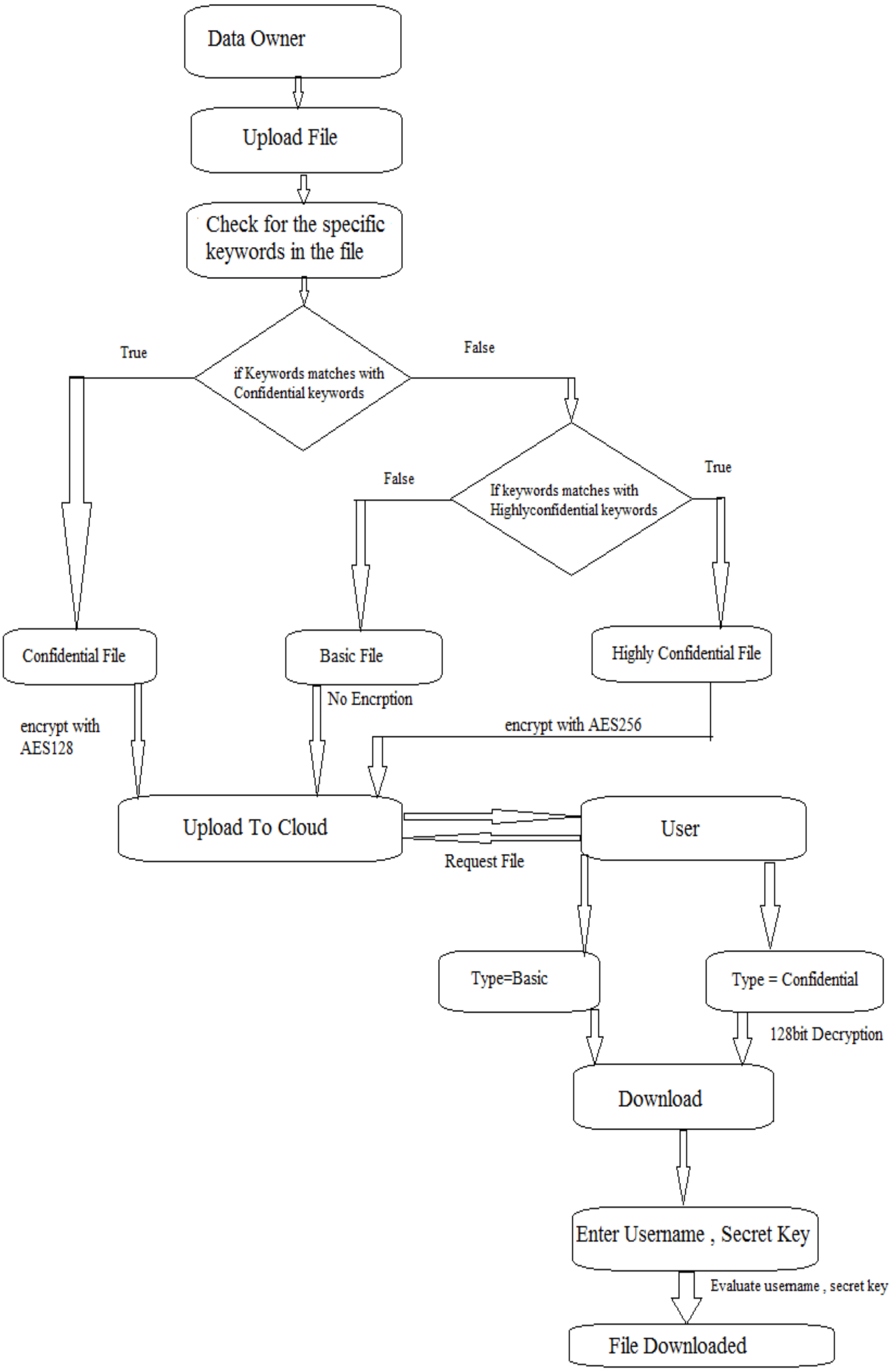


Figure 1: Flowchart for Auto classification of the data.

X ADVANCED ENCRYPTION STANDARD

AES[28] makes use of symmetric encryption algorithm which uses the block size of 128bits and the sizes of the key are 128, 192 or 256 bits respectively. Every round in AES comprises of operations like as SubBytes, ShiftRows, MixColumns, and AddRoundKey and the bytes and keys are of finite field elements. In the year 2001 was considered instead of DES as the technique for encryption by the US government and premeditated of its as standard method for encryption and decryption of the categorized data in a confidential manner. AED does offers many of the significant features, which are like in similar to global ingress, secured type information, free only one of the encryption and also decryption type key which is essential and makes to a convenient installation. Every round in AES does comprises of the four kinds of operation. These can be identified as the SubBytes, ShiftRows, MixColumns and also the AddRoundKey. The bytes of the field elements are of finite and are not just a numericvalue represented for the operation.Generally, the finite field elements are of size 128 and 256. AES makes use of different number of rounds example 10 round for 128bit, 12 for 192bit,14 for 256 bits respectively. Each of them performs the same operations in a round except for the last where it replaces the mixColummns kind operations. The reverse operation of encryption is used to decrypt the data and the relevant symmetric key is utilized to encrypt the text.

Algorithm

```
Procedure
Monitoring files
Set A[] = Users data
repeat
Size = file Size
Switch File Type
Case file in Basic
    starttime= Inittime
    copy(file,Destination)
    endtime=Inittime
EndCase
Case file in confidential
    starttime =Inittime
    AES128(file,Destination)
    endtime=Inittime
EndCase
Case file in Highlyconfidential
    starttime = Inittime
    AES256(file,Destination)
    endtime=Inittime
EndCase
EndSwitch
difference = endtime – starttime
WriteResults(File.Size,starttime,endtime,difference)
Until File in A[] = NULL
```

XI TLS PROTOCOL

TLS is the procedure which does guarantees the secrecy in conveying basically between the users in a network. TLS ensures that there will be none of the interface kind message recognized from the third party. TLS constitute of TLS record type and TLS handshake protocols. The TLS record protocol does offer safer connection. It can be utilized with those of encryption like DES and can be utilized without any of encryption. The authentication between the respective server and the client is facilitated by the TLS type handshake protocol and does permits to decide and make a proper selection of cipher type algorithm and of the cryptographic keys basically before making a exchange of data. This TLS protocol is reckoned on the Netscape SSL 3.0 kind protocol, but inter type operation are not permitted in TLS and the SSL. TLS kind protocol comprises of mechanism which permits TLS enactment in making of backdown to SSL 3.

XII HASH FUNCTION

This can be recognized as the cipher which transforms an arbitrary type of elements of data similar to transforming files of text into a respective hash or the value with the constant size. Later guesstimated hash value is utilized check for the integrity of the original information in the replicas without allowing to get data which are defacto original type. This does mean that when it has value and is actually used for the relative proper reasons, later it is reserved and executed in a convenient phase. Four hashes from SHA-2 comprising 224, 256, 384 or 512 bits.

XIII DETAILS OF SCHEMA

ECCSF constitute of majorly 3 levels of security which is relied on the cloud storage schema. Three levels can be recognized. These are Elementary level, Unrevealing Level and Highly Confidential Level respectively.Elementary Level is used for encrypting generally recognized data which are needed for higher levels of the security which constitute of digital data. Typically, we are considering HTTPS and TLS basically to assure the secrecy in conveying especially between the usersSDSys is recognized a proper solution for this kind of issue by offering a centralized administering layer, which can be made to ingress by the system administrator and is bounded for other respective users.

XIV UNREVEALED LEVEL

It is utilized basically to protect those data which requires the medium level of confidentiality, comprising of individual information. So, we make use of AES128 bit for these types of data.

XV SEVERELY CONFIDENTIAL

It is the highest level and is utilized when the information is highly important to the user which he does not wish to disclose it so such information should be safeguarded utilizing the strong level approaches at possible.

XVI TESTING

Some of the important test cases applied during testing are

Table I: Unit Test Case 1

Test Case ID	Unit Test Case 1
Description	Check whether data is classified
Input	Data selected from the user as basic
Expected output	Insert data into classified level Basic
Actual Result/Remarks	Successful classification of data
Passed(?)	Yes

Table II: Unit Test Case 2

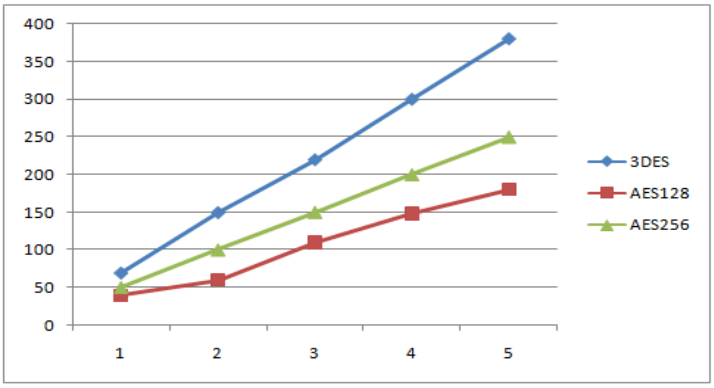
Test Case ID	Unit Test Case 2
Description	Check whether data is classified as Confidential
Input	Data selected from the user as Confidential
Expected output	Insert data is classified as Confidential level
Actual Result/Remarks	Successful classification of data
Passed(?)	Yes

Table III: Unit Test Case 3

Test Case ID	Unit Test Case 3
Description	Check whether data is classified as Highly Confidential
Input	Data selected from the user as Highly Confidential
Expected output	Insert data is classified as Highly Confidential level
Actual Result/Remarks	Successful classification of data
Passed(?)	Yes

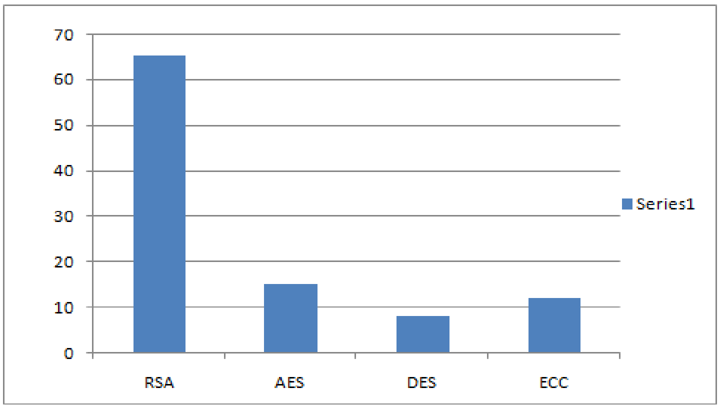
XVII RESULTS

There are various performance charts which can be known in the following.



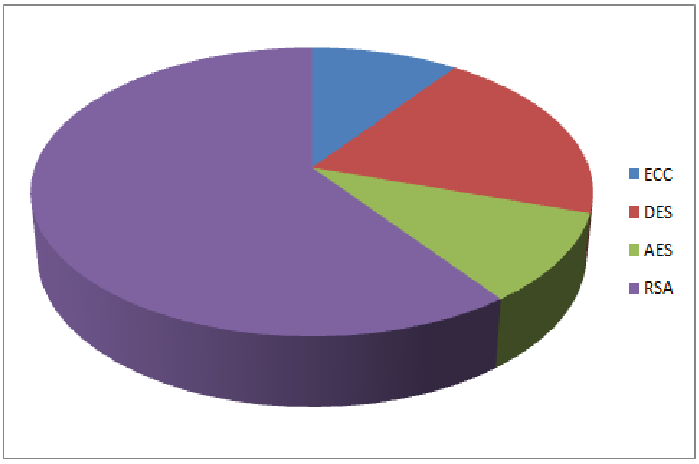
Graph 1: Line graph which represent the performance of encryption algorithms: 3DES, AES128 and AES256

The above graph represents the performance evaluation of proposed work which generates results based on data blocks for the encryptionalgorithms (AES).



Graph 2: Bar chart representing execution time taken by the corresponding algorithms.

The above graph represents time taken to encrypt a particular file by an individual algorithm based on the encryption time taken graphs have been plotted. As per above graph RSA requires higher encryption time and DES requires less encryption time.



Graph 3: Pie Chart representing key size of the encryption algorithms.

The above graph expresses the key sizes of four different algorithms to encrypt data, as per above graph RSA contains large key size, ECC contains smaller key size.

XVIII CONCLUSION

The proposed work suggests an efficacious unrevealing reckoned cloud storage schema which improvises the processing period, assures unrevealing feature and also correctness of information by segregation of the data and applying TLS protocol, AES and SHA algorithm, which is reckoned on the nature of segregated and categorized information. Suggested schema has been considered forward with simulations. The results of the simulations, which exhibits that the selected schema achieves better time of processing while making assurance of the data in unrevealing manner and also observed integrity. We also use self –regulated classification of the data based on the keyword specified, based on the matched classification level encryption is applied.As a part of our future scope weconsider forth in embellishing our schema by auto classification of digital data (audios, videos, images).

REFERENCES

[1] P. Mell and T. Grance, “The nist definition of cloud computing,” ComputerSecurity Division, Information Technology Laboratory, NationalInstitute of Standards and Technology Gaithersburg, Tech. Rep., 2011.

[2] V. Guzhov, K. Bazhenov, S. Ilinykh, and A. Vagizov, “Cloud computingsecurity issues,” in The 2-nd Indo-Russian Joint Workshop on ComputationalIntelligence and Modern Heuristics in Automation and Robotics,2011, pp. 128–133.

[3] F. Ogig?au-Neamt,iu, “Cloud computing security issues,” Journal ofDefense Resources Management (JoDRM), no. 02, pp. 141–148, 2012.

[4] J. Wu, L. Ping, X. Ge, Y. Wang, and J. Fu, “Cloud storage asthe infrastructure of cloud computing,” in Intelligent Computing and Cognitive Informatics (ICICCI), 2010 International Conference on.IEEE, 2010, pp. 380–383.

[5] (2013, sep) Boxcryptor. [Online]. Available: https://www.boxcryptor.com

[6] (2013, sep) Crashplan. [Online]. Available: https://www.code42.com/crashplan/

[7] (2013, sep) Dropbox. [Online]. Available: http://www.dropbox.com

[8] (2013, sep) Mozy. [Online]. Available: http://mozy.com

[9] T. Brindha, R. Shaji, and G. Rajesh, “A survey on the architecturesof data security in cloud storage infrastructure,” Engineering and Technology (IJET), vol. 5, pp. 1108–1114, 2013.

[10] S. Kamara and K. Lauter, “Cryptographic cloud storage,” in FinancialCryptography and Data Security. Springer, 2010, pp. 136–149.

[11] Y. Wei, Z. Jianpeng, Z. Junmao, Z. Wei, and Y. Xinlei, “Design andimplementation of security cloud storage framework,” in Instrumentation,Measurement, Computer, Communication and Control (IMCCC),2012 Second International Conference on. IEEE, 2012, pp. 323–326.

[12] M. Borgmann and M. Waidner, On the security of cloud storageservices. Fraunhofer-Verlag, 2012.

[13] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “Secure RankedKeyword Search over Encrypted Cloud Data,” in Proc. of ICDCS’10, 2010.

[14] S. Yu, C. Wang, K. Ren, and W. Lou, “Achieving Secure, Scalable,and Fine-grained Data Access Control in Cloud Computing,” in Proc.of IEEE INFOCOM, 2010.

[15] W. A. Wayne, “Cloud hooks: Security and privacy issues in cloud computing,” in Proc. of Hawaii International Conference on System Sciences, 2011.

[16] D. Petri, “What You Need to Know about Securing Your Virtual Network,”http://www.petri.co.il/what-you-need-to-know-about-vmwar e-virtualization-security.htm/, 2009.

[17] Texiwill, “Is Network Security the Major Component of Virtualization Security?” 2009.

[18] Y. Jararweh, L. Tawalbeh, F. Ababneh, and F. Dosari, “Resourceefficient mobile computing using cloudlet infrastructure,” in Mobile Ad-hoc and Sensor Networks (MSN), 2013 IEEE Ninth InternationalConference on. IEEE, 2013, pp. 373–377.

[19] M. Quwaider and Y. Jararweh, “Cloudlet-based efficient data collectionin wireless body area networks,” Simulation Modelling Practice andTheory, vol. 50, pp. 57–71, 2015.

[20] Y. Jararweh, L. Tawalbeh, F. Ababneh, A. Khreishah, and F. Dosari,“Scalable cloudlet-based mobile computing model,” Procedia ComputerScience, vol. 34, pp. 434 – 441, 2014, the 9th InternationalConference on Future Networks and Communications (FNC’14)/The 11th International Conference on Mobile Systems and PervasiveComputing (MobiSPC’14)/Affiliated Workshops. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050914009065

[21] L. Tawalbeh, Y. Jararweh, F. Dosariet al., “Large scale cloudletsdeployment for efficient mobile cloud computing,” Journal of Networks,vol. 10, no. 01, pp. 70–76, 2015.

[22] H.-Y. Lin and W.-G. Tzeng, “A secure erasure code-based cloud storagesystem with secure data forwarding,” Parallel and Distributed Systems,IEEE Transactions on, vol. 23, no. 6, pp. 995–1003, 2012.

[23] R. Seiger, S. Gros, and A. Schill, “Seccsie: a secure cloud storageintegrator for enterprises,” in Commerce and Enterprise Computing (CEC), 2011 IEEE 13th Conference on. IEEE, 2011, pp. 252–255.

[24] U. Somani, K. Lakhani, and M. Mundra, “Implementing digital signaturewith rsa encryption algorithm to enhance the data security of cloud in cloud computing,” in Parallel Distributed and Grid Computing(PDGC), 2010 1st International Conference on. IEEE, 2010, pp. 211–216.

[25] Q. Zhang, B. Luo, W. Shi, and A. M. Almoharib, “Cloudsafe: Storingyour digital asset in the cloud-based safe,” Wayne State University,Detroit, USA, Tech. Rep., 2013.

[26] I. M. Khalil, A. Khreishah, and M. Azeem, “Cloud computing security:A survey,” Computers, vol. 3, no. 1, pp. 1–35, 2014.

[27] I. Khalil, A. Khreishah, S. Bouktif, and A. Ahmad, “Security concernsin cloud computing,” in Information Technology: New Generations(ITNG), 2013 Tenth International Conference on, April 2013, pp. 411–416.

[28] Aaron Carroll , Gernot Heiser, An analysis ofpower consumption in a smartphone. Proceedingsof the 2010 USENIX conference on USENIX annualtechnical conference, pages 21-21, June 23-25, 2010.

[29] Kevin D. Bowers, Ari Juels, and Alina Oprea.HAIL: a high-availability and integrity layer forcloud storage. In Proc. of the 16th ACM ConferenceonComputer and Communications Security- CCS’09, pages 187-198, 2009.

[30] Hussam Abu-Libdeh, Lonnie Princehouse, andHakim Weatherspoon. RACS: A case for cloudstorage diversity. In Proc. of the 1st ACM Symposiumon Cloud Computing, pages 229-240, June2010.

AUTHOR PROFILE

DAYANANDA LAL N, BE, MTECH, Research Scholar, Department of Electronics and Communication Engineering ,Dr. MGR Educational and Research Institute University,Chennai,India. ,dayanandlal@gmail.com.

Dr. SENTHIL KUMAR K ME, PhD, Professor, Department of Electronics and Communication Engineering, Dr. MGR Educational and Research Institute University,Chennai,India. ksenthilkumar@drmgrdu.ac.in

GAUTHAM B S, MTECH,Dept. of CSE 2nd Year student, Acharya Institute of Technology, Bangalore, Karnataka, India, gautham4gowda@gmail.com.

NARASIMHA MURTHY.M.S, BE, MTECH, (PhD) Research Scholar, Dayanand Sagar College of Engineering, Bangalore, Karnataka, India. china.mur@gmail.com