Research Article

Data Prevention Technique For Securing The Data

Christy A[a], Katragadda Manoj Narayana[b], Katta Mohana Harshith[c]

[a] Professor, Department of Computer Science and Engineering,Sathyabama Institute of Science and Technology,Chennai,India
[b,c] Student,Department of Computer Science and Engineering,SRM Institute of Science and Technology,Chennai,India

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: The main aim of this paper is to safeguard information through data security and is considered convergence, exploration, infiltration, and footprint. Cloud computing can be a computer-based type of computer that provides numerous PCs and devices looking on the services and data needed. It is a decent issue that completely empowers the interference and management of programmable computer resources. Currently cloud security could be a major issue in building computing within the cloud and so the removal of the cloud computing potency is the winning model for security in the cloud deployments. The current work is a part in implementing cryptography on the cloud computing for higher data security. Here we tend to approach another cryptography security to secure information in the cloud data centres.

Keywords: Data Security, Cryptography, Cloud, Data Privacy, Elliptical Curve Cryptography, Data prevention;

Introduction

Data interference may be a major concern for a commercial enterprise during a recent network. The most goals of the protection domain are to forestall sensitive knowledge from unauthorized organizations and monitor data flow to prevent more security risks. Unacceptable interference has serious issues for long and short-run organizations. Obtaining a structured effort to manage data flow within and out of doors of the organization, forestall unwanted access and transactions from the subtle remains a challenge. The information interference and redress method is a very important analysis drawback, that isn't forever attainable as a result of several factors. Recent news and reports indicate that 50% of information within the sector is part or utterly leaked [1]. The precise details of the leak information and also the source are tough to see. However, there are several channels to leak data interference.

Restrictions are also desecrated by channels that are disproportionately managed by email, prompt messaging and different social network supplements to beat this disadvantage, Data prevention systems are enforced. Less analysis has been introduced to tackle the DP downside. Therefore, there is a need, challenge to vogue and to develop a replacement mechanism of DP with the pliability to seek out it. Galvanized by the DP study field, a survey on information o interference and hindrance practices is bestowed throughout this paper. This paper presents the essential methodology of DP besides recent methods in information interference method. This paper ultimately contributes to the matter and challenges of the recent methods with future work.

1.0 *Data prevention standards*

Many studies are done to explain placement of data prevention within the literature. However, definition of the data outpouring and restriction could be a method of content viewing and protection against misuse [2].The study of Information avoidance is growing. It is a minor problem in obtaining information preventing employer behavioral observation [3]. The authors [4] reviewed the DP methodology and its problems for proper interpretation. The DP process has three stages such as data collection, analysis and stages of corrective action as in Fig.1 Data gathering starts with the user's internet logs and data sources. The data collected is entered into the DP analysis phase, which performs law enforcement, policy enforcement, content verification processes and context. Editing validation removes the sender, source id, information access times, scheduling and quantity that start with that amount of data and so on throughout the process. The view and order of stock of all users may be content manipulated from normal statement and tagging change. The identifier also divides information into those predefined people using those security policies with added tests. Ultimately DP schemes decide the issue by selecting appropriate therapeutic activities such as awareness, prevention, and moreover to complete other activities in the safety program dimensions.

Figure 1 shows a simple deployment of a DP, in which the attributes of avoidance are investigated, monitoring for remedial activities. That data path encapsulates the web data separately from the navigation data. The customer input rights also receive the recorded data starting with the database. The three types are related to the data and are mainly collected in the DP variable. That key 2.0 shows the types of information used for DP. The side-by-side information is that information transmitted by the manifestation of a particular case must be converted within same system by dissimilar networks. The data used are those data, which will then accept clients that are grouped alongside the organization of records or email formats within the applications. That organizational information is not encrypted. This third-party data needs customized help. In addition, it is protected by strict access control.

DP-based content monitors sensitive data with standard diagnostic presentations. With a account number, mobile number and other sensitive information are considered. According to this model, the authors [5] anticipated DP with the standard expression. But this approach is unsuccessful and generates high prices. DP is an agent that has the power to change access to confidential information. The author [6] identified different leakage channels while the data was transmitted. This includes portable Media such as the USB or memory cards. The authors evaluate factors related to sensitive data access and limitations according to audit report. The categories of the DP encrypted strategies, which have different types of Data Restriction methods and maintain Rights Tracking and Identification using Ethical investigations and the like, are the general look-and-see of stock options, etc., which are actually the most prominent ways under DP to use cryptographic claims. Also, for detection, this prevents data starting with unauthorized customers. In identification process, information and behavioral tests for rapid mining require several improvements.

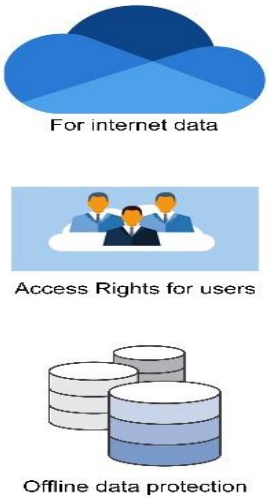


Fig.1 Data trends in Data prevention

1.1 ***Challenges facing during data protection***

- Like alternative ways of data protection and security, DP faces many issues whereas detection and preventing information interference. There are seven major challenges known in previous work [7,20] like the one below. The challenge of beginning information transmission is feasible through multiple channels. If the data transmitted through the appliance you would like are often accessed or protected. However, if a data station isn't a selected application like email, USB and alternative formats are going to be a difficult task.
- The second challenge is conversion, wherever information are often modified and partly charged to users. Generally, several ways realize a pattern of sharp content. Thus, it didn't find fractional and full interference of data conversion with or while not modification.
- The third challenge to search out and supply acceptable access rights to definite users counting on their level of security is an extra challenge. Misguided tips and policies could influence the accuracy of the displaced person.

- The fourth challenge is that the encoding method and the steganography procedure. This procedure will shield the info from unauthorized user, however it's troublesome to research the info content once strict secret writing techniques are used.
- The use of viewpoints preserved the identity and so avoided information protection. However watermarking content isn't reliable for all information sort. And what's additional, the content that's viewed is quickly accessible. This creates several difficult problems.
- The measurability and integration of an oversized domain is not possible. This creates a major drawback. Once network size is giant, then examination policy, observance and access to information is troublesome.
- Finally, the recovery of block information from the log needs an entire supervised learning method. This creates loads of uncertainty and causes many problems in sensitive information. These challenges are usually seen within the literature. These challenges load of analysis has been done and represented within the next section.

2.0 ***Literature survey***

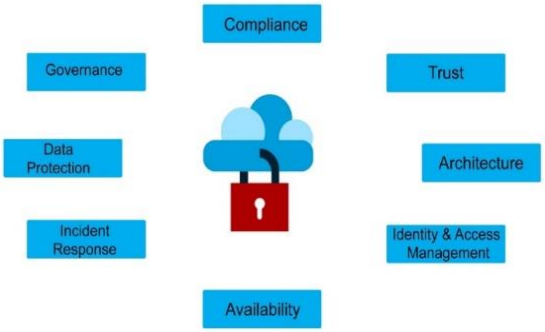
Kunze, m. & Wang, et al. , Tao, J. , Their inquiry report, "Scientific Cloud Computing: Early Meaning and Experience," states that cloud logging benefits consumers, and provides software, hardware and information assets. Some cloud management models:
i) HaaS: The administration may have been recommended for the year 2006 as hardware for service fittings. As well as the rapid development of fixtures virtualization, the mechanization of it, and the increased usage volume, the more expensive customers can buy it fittings - or indeed a full IT / Computer Center - as you go, member management. That Haas is flexible, versatile and reasonable to meet your needs.
ii) SaaS: An explanation of the ease of software as a service product or arrangement is provided with the customer on the Internet, which eliminates prerequisite with the introduction, which further operates those arrangements on the customers nearest computer. SaaS accordingly shifts the client's cerebral pain on product maintenance and further reduces the overhead for product purchases through interest assessment[20].
iii) Toss: With different formats, access to information from different sources, as administrative information, can be accessed by clients with benefits. For example, customers may be able to control remote data, such as initiatives around the neighborhood, such as information on Internet. 2. Er. Rasneet Kaur Chauhan and Er. Sharanjit Singh proposes Cryptographic Calculations that "Introduce Crypto Cloud to Cloud Recording" as follows: i) Information Encryption Terms [DES]ii) Operated Encryption Guidelines [AES]iii) Three - Tess iv) RSA[Rivest, Shamir, Adleman]v) Blowout Opportunity to connect effectively.

Nelson Gonzalez, Charles Myers, Fernando Redaglio, Marcos Simplicio, Mates Nestlundi, Teresa Carvalho, and Magan Poursandi, in their dissertation, point out "What are the results of a cloud-computing investigation of current security concerns?" Data identified with cloud security recognizes policy issues in the region and is comprised of eight different types of compliance, trust, architecture, governance card and more. More access, availability, episode response, information security and more governance.
Christy et al (2019) has proposed a keyword weighting function for document clustering. Each keyword in the sample are clustered based on keyword weighting function. Experimental results were conducted with BBC news collection related to 5 domains and compared with K-Means clustering and Hierarchical clustering algorithms. It is shown that clustering followed by keyword weighting function has improved accuracy[16]. Prayla has come with novel monolithic architecture to design a pattern for biometric service. Here a biometric based passport authentication system is generated wherein the physiological and biometric issues are collected and validated. All the services are treated as micro service and are interfaced with API [17]. Jesudoss et al proposed an authentication model which protects against various number of security attacks. The main aim of the work is provide unique authentication whie satisfying security requirements[18]. M.S.Roobini et al (2019) proposed a method by using algorithm such as ANN, Naïve Bayes, KNN for classification of Diabetes Mellitus [17,18,19].

2.1 ***Cloud Computing***

Cloud computing in historical context is described as a hot point starting from the 2007 claim because of its capabilities, QoS guaranteed record environments, and configurable product management [3] from the demand of multinationals. Cloud computing re-records the web, and this report is basically Tom Cloud exploring the cloud. In the cloud, information can be pushed to a remote area, and accessible depending on the interest. It allows customers to use the request product without having to introduce that document on any machine locally for web connection. The information outsourcing client may start from anywhere and get the necessary data. And the storage room does

not require cerebral pain and can avoid those extra costs associated with software, hardware and data assets as



shown in Fig.2 .

Fig. 2 Cloud computing in historical context

2.2 ***Elliptical Curve Cryptography***

In cloud computing, elliptical curved cryptography can be used as a communication from say, get start with instantaneous public key cryptographic resolutions, Example keys can implement more digital signatures. The different impulses from claiming to use elliptical curves are to reduce life, as well as the minimum critical sizes of the glacier and only the tip of the more workable processes [11]. ECC is a type of open cryptosystem like RSA. What alternative approach for masters about cryptographic computation is that it is likely to be its Snappier-driven breakpoint by influencing differentiation from RSA. A comparable level of protection toward the RSA could be an opportunity to look into the ECC, as well as to a lesser extent. For example, "An RSA's 1024-bit security level is likely to be reduced to the same level of 163-bit protection from demand. Starting from this, the ECC will be ideal for remote communications, such as portable phones, PTAs, cone cards and sensor networks ".

The purpose of the ECC application of the duplication function, which must be computationally traced, is only the tip of an efficient iceberg in RSA exponentiation [12]. A large part of the ecc requirement is considered, as well as those security decisions for remote networks, such as clouds, due to small way scaling and reconfigured computation [13].

Elliptic curve is an attractive property suitable for using cryptography previously done, cloud recording i.e. Get a third reason why its energy is in the same twist in addition to a particular curve that holds the two jumps forward. The basic function of rolling for ecc can be attributed to the existing multiplier, i.e. An extension of a measuring k move toward any parallel p to obtain the additional side of the Q point. [14]. The general scientific expression for an elliptic curve is explained in equation 1.

$$y2 + axy + by = x3 + cx2 + dx + e \quad (1)$$

a, b, c, d and e are the real figures
Let x and y be a set of real numbers.

$$y2 - x8 + dx + e \quad (2)$$

In its simplest form, the elliptic curve can be given by equation (2).

One of the statistical investigations shows that a similar position starting from the protection presented as an enormous modulus towards an RSA-based graph would make the competitors worse for an elaborate very low elliptical curve group. The EC's 163 TouchEnter may safely be the defining one, as are the 1024 odds for RSA to enter. Likewise, the ECC usage is much lower route sizes, which affects faster calculations, higher level of control consumption, saving memory and more transfer speeds. Undoubtedly, cloud logging in this way will provide a significant portion of the cloud with a very secure nature domain about speed and plenty of intangible / intangible assets. EQ, which is integrated with the Cloud, focuses on how to avoid the use of information copying, utility information, and benefits. In addition, ECC placement is very low way sizes, which affects faster calculations, lower power consumption, memory storage and bandwidth. Integrated with in cloud computing, ECC can also provide a secure environment, including the speed and storage of many imperceptible resources.

3.0 ***System Architecture***

In the existing system three step procedure is used. Firstly, Diffie Hellman is used for exchanging keys. Thereafter authentication is performed using digital signature scheme. Finally data is encrypted using AES and then uploaded to the required cloud system. For decryption reverse procedure is implemented.To make the centralised cloud storage secure ECC (Elliptic Curve Cryptography) algorithm is implemented. This approach uses single key for encryption and decryption and complete process takes place at the client side. This methodology performs steps such as:

a. Authentication,
b. Key generation operation,
c. Encryption,
d. Decryption.

The proposed system is a cryptographic algorithm which accepts any kind of data for processing. In addition of that the simulation of the proposed methodology enables a user to send and receive data using the application. The proposed simulation first accepts the data from the user and that it uses the proposed cryptographic algorithm data to manipulate the data into cipher text.

1. Providing the simulation of secure file transfer utility by using hybrid cryptographic algorithm.
2. Designing and implementing the hybrid cryptographic technique in order to reduce the space and time complexity by compressing the data being sent in network.
3. Cross validating the data integrity using the SHA , ECC and RSA function.

After implementation of the proposed cryptographic algorithm the following outcomes are expected.

1. Reduced amount of cipher text
2. Enhanced time and space complexity
3. Authorized with man in middle attack
4. An efficient and robust cryptographic technique.

The system architecture is depicted in Fig. 3.

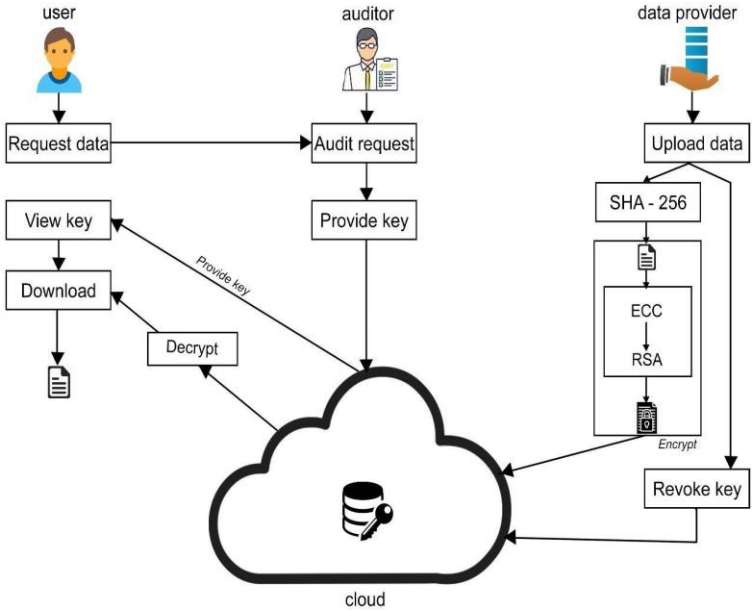


Fig 3. System Architecture

For the purpose of uploading the file, the steps we have performed is depicted in Fig. 4.

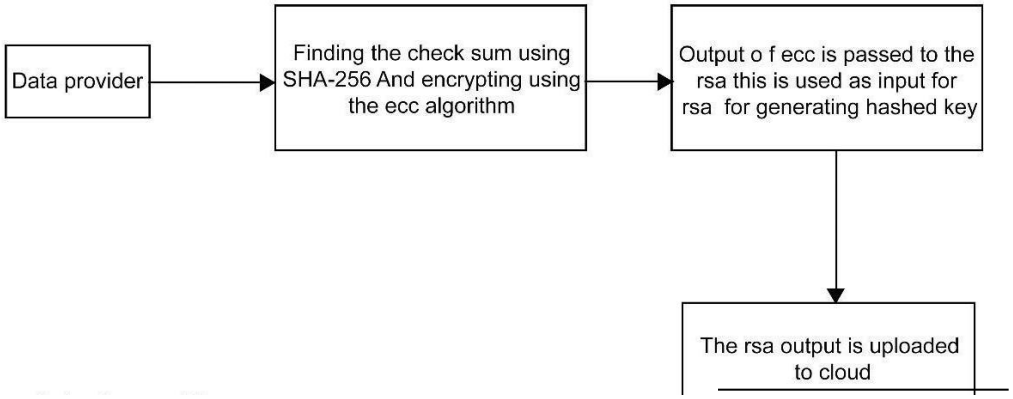


Fig.4 Procedure for uploading a file

In the data provider module, the data given by the data provider will undergo several steps before storing it into cloud. In the first step the data content provided will be passed to the SHA-256 algorithm to get the checksum value and in the next step the data provided will be given to ECC as input and the cipher text generated by Elliptical curve cryptography will be encrypted and the key generated will be encrypted by RSA and the secret key to decrypt the encrypted key will be shared to the user by requesting to the auditor. The dataprovider can revocate the key when he wants to end the previous user not to access the data by generating a new key as shown in Fig. 5.

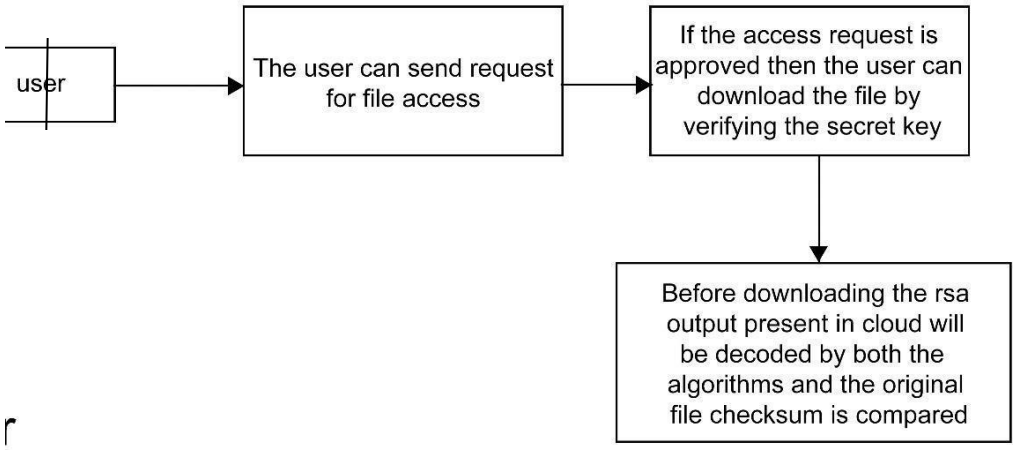


Fig. 5 Data Provider module

The planned HECC message is coded with the bilateral code and also the key over the coded message is encrypted victimisation the uneven cryptography rule. The creator will currently take the information key and can rewrite it victimisation the uneven cryptography rule thereto the recipient‘s public key. The operation forms a key Block. However the advantage of this is often thought it's extremely reckon intensive the dimensions of the information key that we have a tendency to are taking is extremely tiny so the whole method are going to be drained simply a fraction of second. The creator can collect the information block and also the key that was hold on as a block into a

file and transmits the file to the recipient.

STEP 1 :
At first the mastermind and also the recipient of the information should each agree on the parameters for ECC, that is, the domain parameters of the theme. the sector of domain in ECC is outlined by _p‘ within the prime case and also the combine of_c‘ and _d‘ in the binary case. The elliptic curve is outlined by the constants _a‘ and_b‘ employed in its process equation. The elliptical curve can currently be a plane curve that is easy and cover finite field. It consists of points that satisfy the equation at the side of a distinguished purpose at eternity, denoted

STEP 2 :
The Generator G are going to be process the Monogenous cluster that's generated by the only component. There are many logarithm-based protocols that are accessible, which is able to restore the (Zp)x with the elliptical curve. Here we have a tendency to are selecting Elliptic curve Diffie –Hellman key agreement theme supported Diffie playwright scheme is getting used.
For cryptological application G, that is that the smallest plus range n in order that nG=∞, that in most cases is prime. Since n is that the size of a subgroup of E(Fp) from Lagrange's theorem it is shown that the amount h is Associate in Nursing number.

In cryptological applications this range h, that is that the compound, should be tiny and ideally h ought to be one. Unless and till there's assurance that the parameters are generated by the sure party they have to be valid before we have a tendency to use them. The generation of domain parameters isn't typically done by every participant as this method involves computing the number of points on a curve. This method is computing intensive, long and difficult to implement. There are few domain parameters that are already planned by the quality bodies for many common field sizes to beat this reason. These domain parameters that are termed as named curves which might be the federal agency (National Institute of Standards and Technology) curve. The above-named algorithms can generate the curve supported the several fields. However if any user desires to make their own parameters than there are many methods that are accessible and mentioned below which might be wont to generate the curve with the specified set of points on the field:

Firstly, The user will choose the random curve so apply a general point-counting rule like Schoof's algorithm or Schoof–Elkies–Atkin rule, Select a random curve in a very family of curves just like the Koblitz curve is used or choose the amount of points needed so generate a curve with this number.

The projected development is compared to a finite field cryptography (e.g., DSA) that may want 3072-bit public keys and 256-bit personal keys, and number factoring cryptography (e.g., RSA) which needs a 3072-bit values of n, Hence, though the general public secret is giant, the personal key which might be tiny are going to be used for secret writing. With the smaller key size less computation and quicker process are going to be achieved. particularly smaller processors are gift.

4.0 ***Results and Discussion***

The data given by the data provider will undergo several steps before storing it into cloud. In the first step the data content provided will be passed to the SHA-256 algorithm to get the checksum value and in the next step the data provided will be given to ECC as input and the cipher text generated by Elliptical curve cryptography will be encrypted and the key generated will be encrypted by RSA and the secret key to decrypt the encrypted key will be shared to the user by requesting to the auditor. The data provider can revocate the key when he wants to end the previous user not to access the data by generating a new key.

The auditor has the access to check all the user and data provider details registered, user requests for data and the data uploaded to cloud. The secret will be shared to the user and made available to download only after the request approved by the auditor. The user can get the access to the data available in cloud by requesting the auditor for file access upon acceptance the secret key will be shared to the respective user and the user can decrypt the file and download the original data from cloud.

A checksum is a small-sized datum derived from a block of digital data for the purpose of detecting errors that may

have been introduced during its transmission or storage. By themselves, checksums are often used to verify data integrity but are not relied upon to verify data authenticity.

ECC, an elliptic curve is a set of points (x, y), for which it is true that y2=x3+ax+b specified certain selected numbers a and b. Normally the numbers are integers whole numbers, while in principle the system also working with real fractional numbers. Despite what the name propose, the curves does not have an elliptic figure such as a =-4 and b=0.67 provides the elliptic curve with equation y2=x3-4x +0.67. Deploying the application in tomcat server by creating the environment in Amazon web services is depicted in Fig.6.

Fig 6. Deploying the application in tomcat server

The homepage contains registration, data provider, user and auditor tabs as shown in

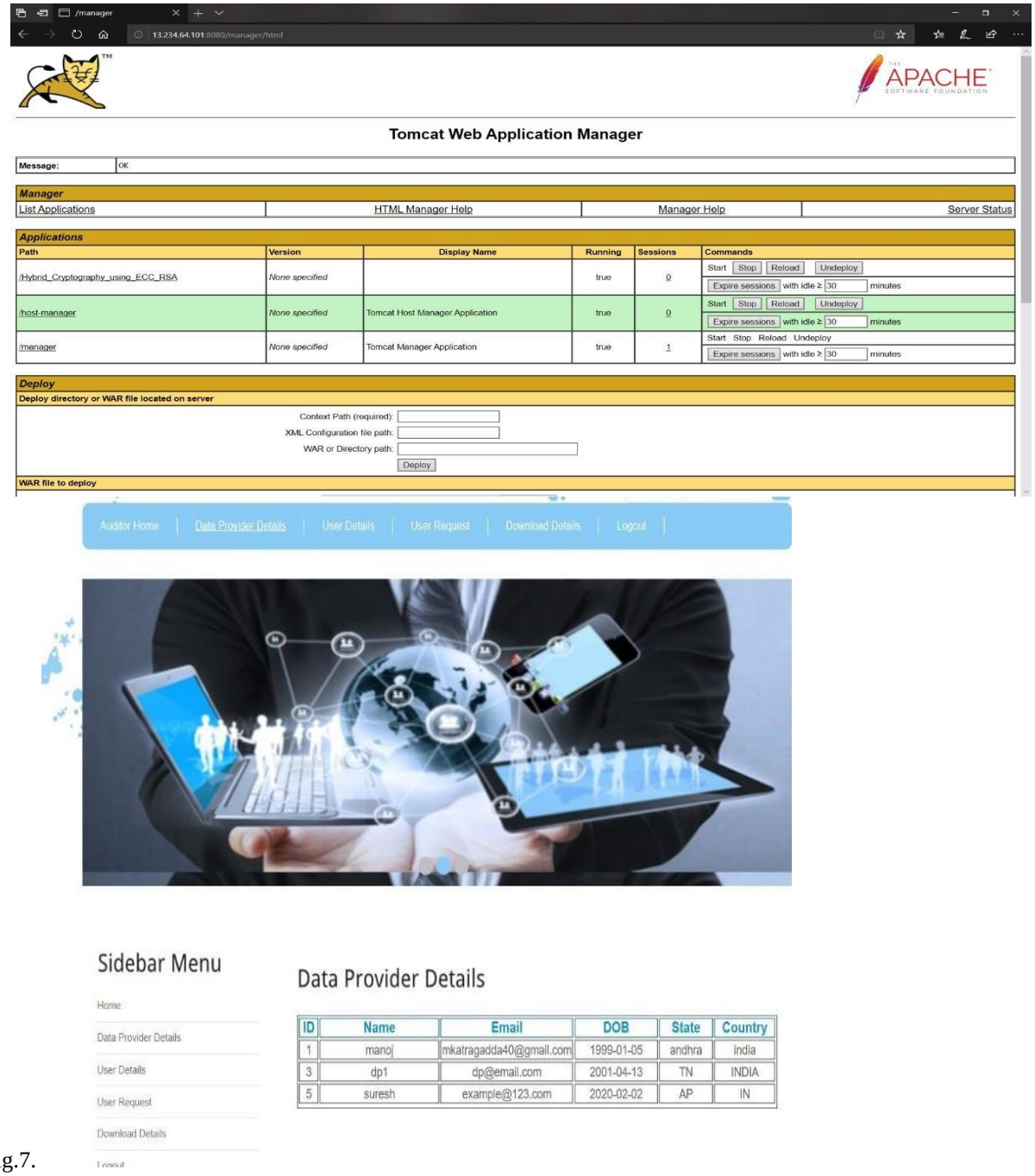


Fig.7.

Fig. 7 Auditor Page

The user window contains the uploaded file details for which the user can send request and file key tab to find the secret key for auditor approved requests and download tab in which the user can download the file by decrypting the file using the secret key.

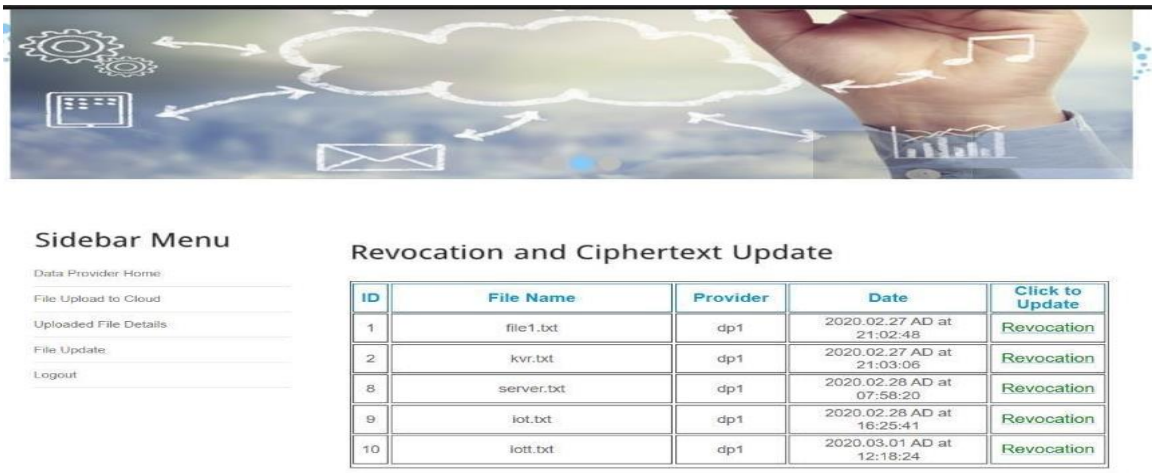
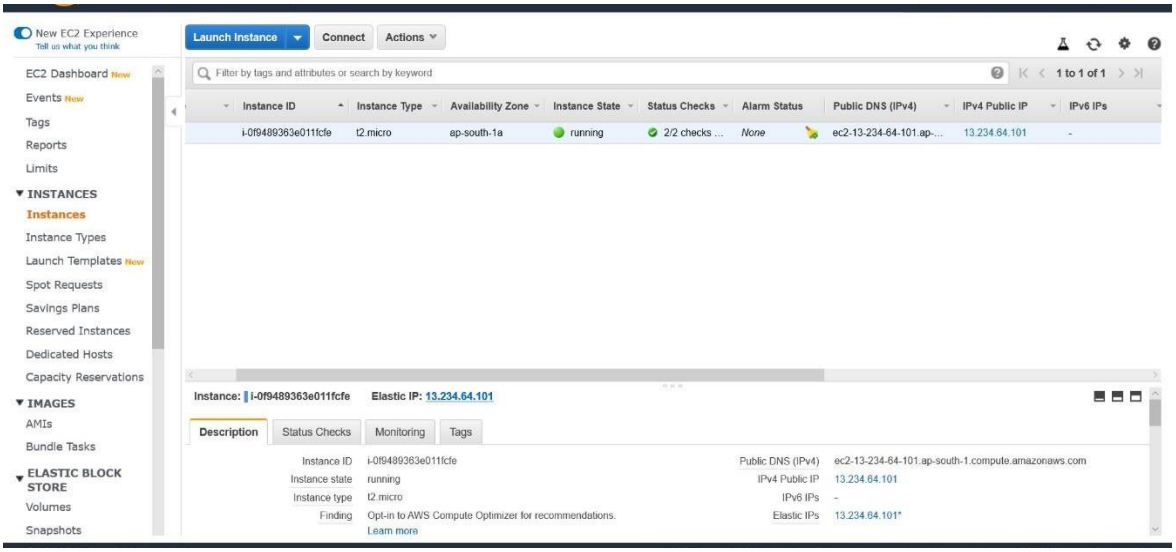


Fig.8 Auditor PageThe user window contains the uploaded file details for which the user can send request and file key tab to find the secret key for auditor approved requests and download tab In which the user can download the file by decrypting the file using the secret key. Key pairs will be generated for the contents to find the matrix. The instance in EC2 to deploy the project in tomcat server is depicted in Fig 9.



5. 0 *Conclusion*

Elliptic Curve Cryptography provides greater security and more efficient performance than the first-generation public key techniques like RSA now in use. As vendors look to upgrade their systems they should seriously consider the elliptic curve alternative for the computational and bandwidth advantages they offer at comparable security. Although ECC's security has not been completely evaluated, it is expected to come into widespread use in various fields in the future. After comparing the RSA and ECC ciphers, the ECC has proved to involve much less overheads compared to RSA. The ECC has many advantages due to its ability to provide the same level of security as RSA yet using shorter keys. However, its disadvantage which may even hide its attractiveness is its lack of maturity, as

mathematicians believed that enough research has not yet been done in ECC. The future of ECC looks brighter than RSA as today's applications (smart cards, pagers, and cellular telephones etc.) cannot afford the overheads introduced by RSA. At least, in today's small computing devices ECC can be used for encryption and decryption as it requires smaller key sizes and has lesser computing complexity as compared to RSA. Thus, ECC makes it an ideal choice for portable, mobile and low power applications and their integration with cloud services. This work compares the time taken by the two algorithms for key generation and encryption. The importance of this work is to use ECC algorithm in cloud storage which has better security services. This work can be extended to compare ECC with other algorithms used for digital signatures, key exchanges as well as to provide the data integrity.

References

1. Data loss db. Data loss statistics. Retrieved from ⟨http://datalossdb.org/⟩; 2015
2. Mogull R. Understanding and selecting a dataloss prevention solution. Retrieved from 2010
3. Boehmer, Wolfgang. "Analyzing Human Behavior Using Case-Based Reasoning with the Help of Forensic Questions." In Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on, pp. 1189-1194. IEEE, 2010.
4. Shabtai, Asaf, Yuval Elovici, and LiorRokach. "A survey of Data Prevention and prevention solutions". Springer Science & Business Media, 2012.
5. Yu, Fang, Zhifeng Chen, YanleiDiao, T. V. Lakshman, and Randy H. Katz. "Fast and memory-efficient regular expression matching for deep packet inspection." In Architecture for Networking and Communications systems, 2006. ANCS 2006. ACM/IEEE Symposium on, pp. 93-102. IEEE, 2006.
6. Shweta Sharma, Bharat Bhushan, Shalini Sharma -" Improvising Information Security in Cloud Computing Environment"- International Journal of Computer Applications (0975 – 8887) Volume 86 – No 16, January 2014.
7. Status Project [URL]. http://www.acis.ufl.edu/vws/, access on June 2008.
8. Christy, A., Gandhi, G.M., Vaithyasubramanian, S. (2019)," Clustering of text documents with keyword weighting function", International Journal of Intelligent Enterprise, Vol. 6, Issue.1, DOI: 10.1504/IJIE.2019.100029
9. Dr.R.Shanmugalakshmi, M.Prabu – "Research Issues on Elliptic Curve Cryptography and Its applications"- IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.6, June 2009.
10. Wang, H., Sheng, B. and Li, Q. (2006) 'Elliptic curve cryptographybased access control in sensor
11. networks', Int. J. Security and Networks,Vol. 1, Nos. 3/4, pp.127–137.
12. Ms Bhavana Sharma, B.P.I.T., Rohini, Delhi-"security architecture of cloud computing based on elliptic curve cryptography (ecc)" ICETEM 2013.
13. Wikipedia, the free encyclopedia of Cloud Computing
14. Prayla Shyry S. (2020) Biometric-Based Three-Tier Microservice Architecture for Mitigating the Fraudulent Behaviour. In: Kundu S., Acharya U., De C., Mukherjee S. (eds) Proceedings of the 2nd International Conference on Communication, Devices and Computing. Lecture Notes in Electrical Engineering, vol 602. Springer, Singapore".
15. Jesudoss A. and Subramaniam N.P., "EPBAS: Securing Cloud-Based Healthcare Information Systems using Enhanced Password-Based Authentication Scheme", Asian Journal of Information Technology, Vol. 15, Issue 14, 2016, pp. 2457-2463.
16. Joseph Manoj, R., Anto Praveena, M.D., Anvesh, M., Pujith, M. "Secured User Behaviour Based Access Framework for Web Service", IOP Conference Series: Materials Science and Engineering, 2019, Vol.590, pp.1-12
17. M.S.Roobini,DrM.Lakshmi,(2019),"Classification of Diabetes Mellitus using Soft Computing and Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering,ISSN: 2278-3075, Volume-8, Issue- 6S4
18. Nagarajan, G., Minu, R. I., & Devi, A. J. (2020). Optimal Nonparametric Bayesian Model-Based Multimodal BoVW Creation Using Multilayer pLSA. Circuits, Systems, and Signal Processing, 39(2), 1123-1132.
19. Nagarajan, G., & Minu, R. I. (2018). Wireless soil monitoring sensor for sprinkler irrigation automation system. Wireless Personal Communications, 98(2), 1835-1851.

20. Nagarajan, G., & Thyagharajan, K. K. (2012). A machine learning technique for semantic search engine. Procedia engineering, 38, 2164-2171.
21. Nagarajan, G., R. I. Minu, V. Vedanarayanan, SD Sundersingh Jebaseelan, and K. Vasanth. "CIMTEL-mining algorithm for big data in telecommunication." International Journal of Engineering and Technology (IJET) 7, no. 5 (2015): 1709-1715.