Privacy Retaining Data Encryption Method for Big Data in Cloud Computing

Meghashree D. P.[1] Nagamani D. R.[2]
[1]Student [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]Bangalore Institute of Technology, Bangalore-560004, India

Abstract— When big data applications are increasing, privacy becomes major issue. The implementation of these kind emerging technologies has the advantages like improved, changed service models & improve the applications performances in different perspectives. The increasing amount of data also leading a several challenges in practice. The time required for the data encryption is the serious issues during the transmission and processing of the data. Several applications leave data encryptions to reach increasing performance level including privacy. In the proposed approach concentrate is on privacy and propose encrypting data model, called Dynamic Data Encryption Strategy. Our proposed approach is aimed to encrypting the data selectively on the basis of privacy value. Designed approach is to increase the privacy protection using a selective encryption method involving required execution time.

Key words: Privacy, Cloud Computing, Data Encryption

I. INTRODUCTION

A. Cloud - a sensible approach, expertise, price advantages & potential rework, information center from intensive originated, different cost setting. Cloud relies on very basic reusable construct of IT principle capabilities. Cloud is defined as the; "Atmosphere, of computing resources and managed infrastructure work out capable to finish applications of client.

Introducing cloud techniques empowered varied applications to people recent years. Human's involvement within cloud associated wired less affiliation loops become alternation to data retrieval derivation from perceptive behaviors of human's interactivities within numerous social networks & apps. Despite several advantages of exploitation cloud, some considerations in protective knowledge privacy of owners throughout communications within social networks & apps.

The drawbacks is that the issues like contradictions in knowledge transmission potency, protection and to solve problem, tend to propose completely unique strategy that encrypting data so maximize degree of encrypted data, beneath the specified temporal order constraints. Cloud is associate internal connective mobile users platform that support sharing of data among several parties across different kind infrastructure. Communications between users typically carry data like personal information through spread channels, as well as sites of social network & infrastructure data secure management, storage, transmissions- these crucial kind aspects are included within previous researches.

A. *Cloud Advantages*

Enterprises wants align applications, thus on exploit, design models cloud offers several everyday advantages below listed:

1) Price cuts

There are many reasons why cloud technology is falling into lower fees. The charging model is paid according to use; no infrastructure is purchased, so renovation is decreased. In historic calculations, initial costs and retirement prices have been very rich.

2) Increase Storage Space

Now, the massive infrastructure provided by cloud vendors, storing and retaining huge amounts of data may additionally have grown to be a truth. As the cloud will enlarge dynamically, surprising peaks of employment have additionally been correctly and speedy controlled.

3) Flexibility

This is an essential function. As organizations ought to adapt to a dynamic enterprise surroundings, the velocity of handing over deliveries is essential, and cloud computing emphasizes that the applications to be received can pressure applications quicker and more fast, and dynamic business situations and the rate of transport are important. Cloud computing specializes in obtaining applications by means of sacrificing the most important and perfect building blocks important to quickly promote applications.

B. *Cloud Models*

The services supplied by cloud carriers can be divided into 3 categories.

1) Software as a Service (SAAS)

In the SAAS model, whole software is supplied to clients on call for. One instance of this service runs at the cloud and serves multiple end users. On the consumer facet, no up-front funding inside the license of the server or software program is needed, and for the provider, the cost is lower due to the fact simplest one software wishes to be hosted and maintained. Now, software program as a service is furnished through Google, Salesforce, Microsoft, Zoho and others.

2) Platform as a Provider (PAAS)

In platform-as-a-service, the software program layer and development surroundings encapsulated, supplied as offerings in order that different excessive-degree services may be built. Customers are loose to build their own programs that run at the issuer's infrastructure. To meet the necessities of software manageability and scalability, PaaS companies provide predefined OS and application server collections, inclusive of LAMP systems (Linux, Apache, MySql, and PHP), restrained J2EE, Ruby, and others. Google App Engine, Force.Com, & so on. Are some popular systems as examples of services.

3) Infrastructure as a Service (IAAS)

This gives primary garage and computing capabilities a standardized service at the community. Servers & storage structures, community gadgets, records middle spaces, etc. Are aggregated and used to address workloads. Customers regularly set up their personal software at the infrastructure. Examples - Amazon, GoGrid, 3 Tera, etc.

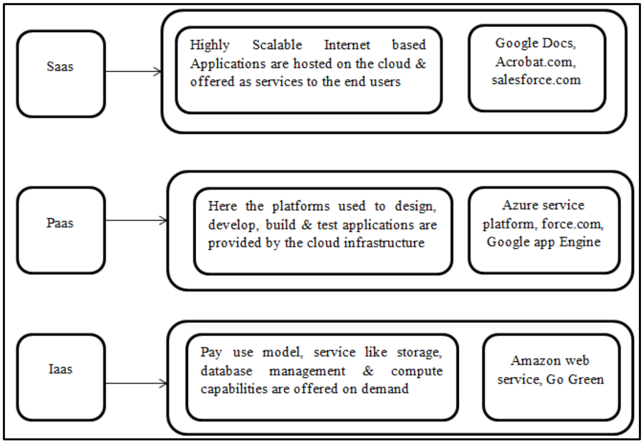
Saas
Highly Scalable Internet based Applications are hosted on the cloud & offered as services to the end users
Google Docs, Acrobat.com, salesforce.com
Paas
Here the platforms used to design, develop, build & test applications are provided by the cloud infrastructure
Azure service platform, force.com, Google app Engine
Iaas
Pay use model, service like storage, database management & compute capabilities are offered on demand
Amazon web service, Go Green

Fig. 1.1: Models of Cloud

C. *Cloud Types*

Enterprise selects application packages to install on public, non-public or hybrid clouds. Integrators of cloud play vital role in figuring out a well-prepared cloud course to organization.

1) *Public Cloud*

They owned & operated with the aid of third party events. They provide customers with terrific economic scale. The value of infrastructure is dispersed among all users, presenting each pattern with an appealing low-cost "pay as you pass" version. All clients share the identical infrastructure pool with much less configuration, safety protection and availability variations. These are managed and supported with the aid of cloud carriers. One of the essential advantages of public clouds is that they may be larger than organization clouds, if you want to scale seamlessly on demand.

2) *Private Cloud*

Private clouds might be mainly constructed for a single agency. They are designed to cope with information security troubles and offer greater manipulate, normally now not inside the public cloud. There are two modifications to the non-public cloud:

a) Internal Deployment of Personal Clouds

Internal non-public clouds (additionally called internal clouds) are hosted of their own information facilities. This version presents greater standardized procedures and protection, however is constrained in size and scalability. The IT branch additionally wishes to endure capital and operating fees for bodily assets. This is first-rate acceptable for programs that require whole manage and configurability of infrastructure and safety.

b) Externally Hosted Personal Cloud

This private cloud is hosted externally through the cloud company, and the vendor implements a non-public cloud by using completely shielding privateers. For agencies that do not like public clouds due to the fact they percentage physical resources that is the most appropriate cloud.

3) *Hybrid Cloud*

This cloud combines public and private cloud models. With hybrid cloud, carrier companies can absolutely or partially leverage cloud carriers to maximize computing flexibility. The hybrid cloud can offer on-call for, externally supplied scale. The ability to leverage the resources of the public cloud to expand the private cloud may be used to manipulate the surprising surge in any workload.

D. *Existing System*

In the Existing system the privacy issue caused because unencrypted information transmission. Considering the appropriate performance level, several applications leaves cipher texts in mobile cloud information transmission. This will be leading to leakage of privacy issue because plain texts are unchallenging to adversaries for information capture in different ways. Issue of privacy exigent, as result, this faces contradiction in protection & performance is sometimes attaching to time constraint.

1) *Disadvantages of Existing System*

- Less efficiency
- Security is less
- Speed of transmission is less.

E. *Proposed System*

In the proposed system we tend to consider privacy, propose a unique encryption that named has Dynamic Data Encryption Strategy. Approach proposed encrypting data selectively; use classifying privacy strategies underneath constraints of time. It aims to increase protection of privacy using a strategy like encryption by selection among specified time execution needs.

1) *Advantages of the Proposed System*

- Maximize protection in privacy by employing selective strategy of encryption among the desired time needs.
- Efficiency and Timing constraints.

II. Literature Survey

Sherif T [1] Amin explains the DNA encryption algorithm in his paper. The basic idea involved in DNA encryption, consisting of storage ability & parallelism, to perform traditional cryptographic algorithms, here the plaintext message is transformed to DNA sequence four binary nucleotides represents the binary octet of a single plain text character.

Cong Wang [2] explains that cloud computing has become one of the next generation information technique for the enterprises because of its unprecedented advantage in IT. It provides on-call for, self-provider, ubiquitous network get entry to an independent region aid pool, rapid useful resource scalability & use based cost charge and transformation risk. The fundamental idea is shifting the data are being outsourced to cloud.

Matteo Maffei [3] explains about cloud become a cornerstone for several information technology infrastructure and provides a solution to backing up, transforming large amount data uploading & synchronization. Data of user is direct control to service providers of cloud, it leads to safety & privacy concern these are associated with the integrity of outsourced facts. The Unintended / intentional disclosure of private data & user activity evaluation, and so forth. If cloud issuer is trusted, the user might get entry to the uploaded file may be malicious and misbehaves.

Siyuan Lin [4] explained that data/information needs to secure & effective, including information privacy and device overall performance. Problem is about allotted information sharing and privacy, safety. Given that statistics demanders request information from disbursed data companies, the purpose is enable statistics demanders to get right of entry to allotted records without understanding the privacy of any person issuer. This trouble is challenged by problems: how to transmit records competently and securely; and how to handle data effectively.

Zhang Yuan [5] explained that many schemes are planned to permit 3rd-party to apply user's key to verify the integrity of the information. Most of those plans have robust assumptions: the auditors are sincere and dependable, and consequently liable to the maliciousness of the auditors. SCLPV is the first work that simultaneously helps much less public review and forestalls malicious auditors from verifying the integrity of outsourced information within the CPSS. Compared with the high-quality integrity verification scheme, malicious auditors can withstand malicious audits. The conversation cost between the auditor and SCLPV cloud server has nothing to do with the size of the processed data; on the equal time, the auditor of SCLPV does no longer want to manipulate the certificate.

III. SYSTEM ARCHITECTURE

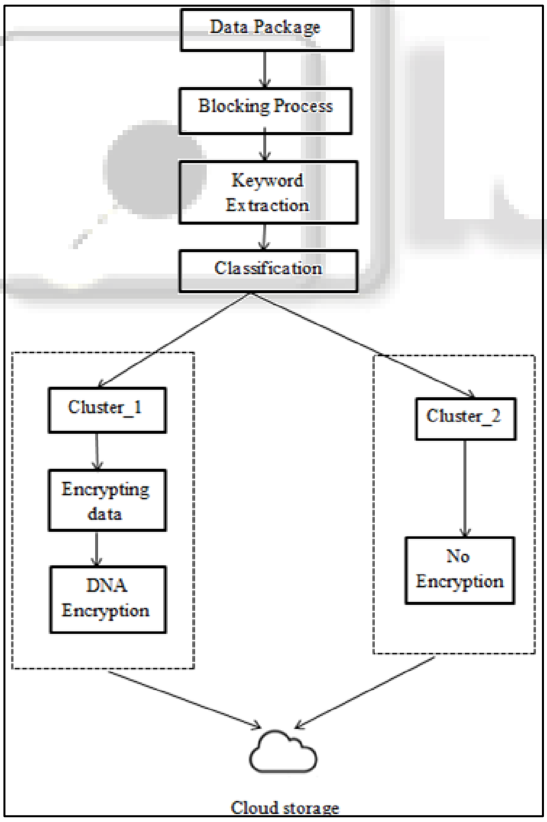

Fig. 4.2: Architecture of System

The figure 4.2 shows the System architecture; the user upload, file selected to cloud while uploading file is divided into blocks, each block key word is extracted. Key word is extracted based on space between the words. These words are formed into array list. Each word in the array list is classified based on cluster _ id mentioned in the m _ terms file in the database. If the word is belongs to the cluster_1 these words are selected for encryption, after encryption these words are stored on to the cloud remaining words are send as it is to the cloud.

IV. METHODOLOGY

A. *Module Design & Description*

1) Admin Module

- In this module admin will be able to login by giving the credentials.
- Admin can upload the file to training data sets by sensitive or non-sensitive based upon the keywords.
- Admin can see the users.
- Admin can be able to delete the user.

2) Training Data Sets Module

- In this module, admin or user can upload their files for training their data sets for keywords based on the sensitive or non-sensitive.

3) User module:

- User registration needs to be done.
- User can be able to login based upon the user id and passwords credentials.
- User can also upload their files. Once the file is uploaded, preprocessing occurs; based on the sensitive and non-sensitive the file will be encrypted or without encryption will be stored in the cloud.

4) De duplication module:

- In this module, once the file is uploaded, the redundant files are checked with it. If the file is repeated then this module will not upload the file to cloud, instead of that, it makes the file instance is two.

B. *Techniques*

1) Splitting Chunk technique

In this technique text file uploading divided into number of chunks on the basis of packet size.

2) Technique of Hashing

Number of chunks divided is converted using MD5 algorithm into hash code.

3) Checking Process for De duplication

Here divided blocks uploaded to cloud, de duplication of blocks is checked before uploading to cloud.

4) Merging Process of Chunk

Here, downloaded blocks from cloud merged and get back original file finally.

C. *DNA Encryption Algorithm*

[This algorithm involves search technique to find & return position of a quadruple DNA sequence represented by plaintext character binary octet.]

1) Inputs

PTC [characters in plaintext], RF [binary random files], RAND [G], DNA SEQ [AAACAGATCACCCGCTGAGCGGGTTATCTGTT], Binary numbers of DNA sequences.

2) Algorithm

1) Step 1. PTC →ASCII [PTC];// In plain textual content file, every character is changed by way of its ASCII code:
2) Step 2. ASCII [PTC]→BF [ASCII];// Convert ASCII values to binary values [0 and 1]: ASCII [PTC]→BF [ASCII];

3) Step 3. ASCII [PTC] →DNA SEQ //Replace the ASCII code with its DNA collection: ASCII [PTC] →DNA SEQ;
4) Step 4. DNA SEQ→ALT [DNA SEQ] //Convert DNA SEQ to Alternative DNA SEQ:
5) Step 5. From a random function in a binary document represented as unmarried-stranded DNA collection, search for a quadruple DNA collection representing pure text characters at a time. This "chasing sequence" has the same series as the apparent textual content individual ASCII code.
6) Step 6. RL→RND [G];// The sequential search starts from the random function RL
7) Step7. If DNA [PTC] = DNA [SEQ] //If the suitable mode is observed, its position might be recorded within the pointer (PNTR) or region output document. Begin looking series SEQ beginning from position RL. This output document is referred to as encrypted textual content but is inaccurate. However, the report carries handiest tips to the octet places located. PNTR takes place;
8) Step8. Repeat the equal steps for all other characters starting at step 2;

– Output

PNTR, a pointer to the placement of the discovered quad-nuclear DNA nucleotide collection representing the binary octet.
Stop

– Tracing of Algorithm

Word=bank ASCII value for b is 98
Value→98
Binary→ 1100010
Binary Full→ 01100010
Output Phase1: TCAC
Output Phase2: AGCG
First 2: AG
Second 2: CG
Test a>>>: AG
Test b>>>:3
Output1:3
Test a>>>: CG
Test b>>>: 7
Output1:7
Test a>>>: AG
Test b>>>:3
Test a>>>: CG
Test b>>>: 7
After Encryption→0307

D. DNA Decryption Algorithm

1) Input

PTC [plaintext characters], RF [random binary files], RND [G], DNA SEQ [AAACAGATCACCCGCTGAGCGGGTTATCTGTT], Binary numbers of DNA sequences.

2) Algorithm

1) Step 1. DEC→ALT [DNA SEQ] //Convert the decrypted word to a DNA candidate SEQ, Start from a random function in a binary document inside the shape of a single-stranded DNA series, and search for a quadruple DNA collection in simple textual content characters.
2) Step 2. RL→RAND [G]; //The sequential search starts from the random position RL.
3) Step 3. ALT [DNA SEQ] →DNA SEQ; //Alternate DNA SEQ to DNA SEQ:
4) Step 4. Then convert the DNA SEQ to its corresponding binary quantity: DNA SEQ→binary No. // A-00
   C-01
   T-10
   G-11
5) Step 5. Binary No translates to its equal hexadecimal price
6) Step 6. Hexadecimal decimal number
7) Step 7. Decimal == ASCII [PTC]

– Output: Decrypted character

Stop

E. Tracing of DNA Decryption

Display 0
1: 0
2: 3
3: 0
4: 7
a→3
b→7
a→ AG
b→ CG
Y→TCAC
zBinary→ 01100010
Enter the Binary number:
Hexa decimal: 62
zBinary Byte→ 62
Decimal: 98
Decimal=98
ASCII: 98
-----b
Stop

V. RESULTS

In this project whatever the data need to encrypt that only encrypted other words are send as it is, it improves the security and reduces the total time to encrypt

VI. CONCLUSION & FUTURE WORK

The proposed work is focus on data's privacy issues & involves implementation practically in cloud computing. The approach proposed, D2ES designed to increase privacy protection efficiency by encrypting data selectively according to, vale of the data privacy. Future work is to increase the size of file to upload and to have more sensitive word database to for the encryption purpose.

REFERENCES

[1] Cong Wang, "Privacy Preserving Public Auditing for Secure Cloud Storage", IEEE Transactions on Computers, Vol. 62, No. 2, pp. 362-375, Dec 2015.

[2] Siyan Lin, "SMC: A Practical Schema for Privacy-Preserved Data sharing over Distributed Data streams", IEEE Transactions on Big Data, Vol. 1, No. 2, pp. 68-81, Apr 2015.

[3] Yuan Zhang, "SCLPV: Secure Certificate less Public Verification for Cloud-Based Cyber-Physical-Social Systems against Malicious Auditors", IEEE Transactions on Computational Social Systems, Vol. 2, No. 4, pp. 159-170, Dec 2015.

[4] Sherif T. Amin, "A DNA-based Implementation of YAEA Encryption Algorithm"

[5] Matteo Maffei, "Privacy and Access Control for Outsourced Personal Records", IEEE Symposium on Security and Privacy, Vol. 2, No. 4, pp. 159-170, Dec 2015.