International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 1, Issue 10, October 2014)

Privacy Preserving Based Cloud Storage System

V.Shravani
M.Tech Scholar
Aurora's Technological & Research Institute
Uppal, Hyderabad.

T.V.Ramanamma
Sr.Asst.Professor
Aurora's Technological & Research Institute
Uppal, Hyderabad.

Abstract: Cloud computing provides huge computation power and storage capability that alter users to deploy computation and data-intensive applications while not infrastructure investment on the process of such applications, an oversized volume of intermediate knowledge sets are going to be generated , and sometimes hold on to avoid wasting the value of re computing them. However, protective the privacy of intermediate knowledge sets becomes a difficult drawbacks as a result of adversaries could recover privacy-sensitive data by analyzing multiple intermediate knowledge sets. Encrypting all knowledge sets in cloud is wide adopted in existing approaches to deal with this challenge. However we tend to argue that encrypting all intermediate knowledge sets square measure neither economical nor efficient as a result of it's terribly time intense and dear for knowledge intensive applications to encrypt or decrypt data sets where as play acting any operation on them. During this paper, we tend to propose a complete unique bound privacy outflow constraint based approach to spot that intermediate knowledge sets ought to be encrypted and that don't, so privacy preserving value may be saved where as the privacy needs of information holders will still be happy. Analysis results demonstrate that the privacy preserving value of intermediate knowledge sets may be considerably reduced with our approach over existing ones wherever all knowledge sets square measure encrypted.

Keywords: - Cloud computing, data storage privacy, privacy preserving, intermediate data set, privacy upper bound

1. INTRODUCTION:

We think about applications wherever the first sensitive knowledge can't be is composed. Perturbation may be a terribly helpful technique wherever the information is changed and created "less sensitive" before being two-handed to agents. One will add random noise to bound attributes, or one will replace actual values by ranges. However, n some cases it's vital to not alter the first distributor's knowledge. For instance, if Associate in nursing outsourcer is doing our payroll, he should have the precise earnings and client checking account numbers. If medical researchers are going to be treating patients (as hostile merely computing statistics), they will would like correct knowledge for the patients. Historically, discharge detection is handled by watermarking. e.g., a singular code is embedded in every distributed copy. If that replicate is later discovered within the hands of Associate in nursing unauthorized party, the informant are often known. Watermarks are often terribly helpful in some cases, but again, involve some modification of the first knowledge. Moreover, watermarks will typically be destroyed if the information recipient is malicious. knowledge discharge are often detailed as in once a knowledge a know- ledge an information} distributor has given sensitive data to a group of purportedly trustworthy agents and a few of the information is leaked and located in associate in nursing enterprise knowledge leak may be a shivery proposition. Security practitioners have continually had to trot out knowledge discharge

problems that arise from numerous ways that like email, and different net channels. Just in case of knowledge discharge from trustworthy agents, the distributor should assess the chance that the leaked knowledge came from one or additional agents.

This can be done by employing a system which may establish those parties World Health Organization area unit guilty for such discharge even once knowledge is altered. For this the system will use knowledge allocation are also can inject realistic however fake knowledge record to enhance identification of discharge. More- over, knowledge also can be leaked from inside a corporation through mails. There's conjointly a requirement to filter these e-mails. This will be done by interference emails that contain pictures, videos or sensitive knowledge for a corporation. Principle utilized in e-mail filtering is we tend to classify e-mail primarily based the finger prints of message bodies, the white and black lists of email addresses and also the words specific to spam.

2. PREVIOUS SYSTEM

In Existing system, watermarking technique is employed to spot the guilty agents World Health Organization area unit unseaworthy the sensitive info or information. Watermarking is one in all previous techniques that contain a novel code. This distinctive code is embedded in every copy that is then distributed to the purchasers by the user.

LIMITATIONS:

However, in some cases it's vital to not alter the initial distributor's information.
Traditionally, cloud discharge detection is handled by watermarking, e.g., a novel code is embedded in every distributed copy. If that duplicate is later discovered within the hands of associate degree unauthorized party, the informant will be known. Water marks will be terribly helpful in some cases, but again, involve some modification

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 1, Issue 10, October 2014)

of the initial information. Watermarks will generally be destroyed if the info recipient is malicious.

3. PROPOSED SYSTEM

The main aim of the projected system is to seek out out once and UN agency has leaked the sensitive information. Within the projected system; it's getting to implement the thought of "Fake Objects". currently if suppose the director of the corporate needs to share some sensitive information (records) with shoppers of his company however he doesn't wish his information to be leaked anyplace in between. Therefore before causation the sensitive information to the shoppers what the projected system can do is it'll add faux objects (record) in information which can precisely appear as if original information. The shopper is going to be unaware of those faux objects.

The system goes to use totally different modules for adding faux objects and or detection faux objects. additionally implementation thought of email filtering module during which if agent tries to e-mail the sensitive information, the mail are going to be sent in Associate in Nursing secret writing format by exploitation cryptography. And also the information's keep within the data set are going to be in encrypted format. If the agent tries to send the info to shopper suggests that the info are going to be send as encrypted data's, At same time the owner of the organization additionally can get identical mail that the agent leaks. However agents don't understand that, however the owner of the corporate will read the first information that is leaked.

ADVANTAGES

• We additionally gift algorithms for distributing objects to agents, in an exceedingly manner that improves our possibilities of characteristic a source.
• Finally, we tend to additionally take into account the choice of adding "fake" objects to the distributed set. Such objects don't correspond to real entities however seem.
• This approach saves the time as a result of we tend to area unit progressing to implement this method solely within the middle knowledge set.

4. MODULE DESCRIPTIONS

User Authentication:

In this module, user registration process is done by the administrator. Here every user will give their details for registration. User details are encrypted by using AES (Advanced Encryption Standard) Algorithm. Authenticated person only can register in this process. Administrator will generate a username and password for every user. The user can use that username and password for login process.

Fake Object Generation:

In this module, the owner of the data or information will add some fake objects (record) in database which will exactly look like original data. The client will be unaware of these fake objects. Only the owner of the data knows where and how many fake objects inserted into original data.

E-Random Implementation:

E-optimal solution

- O (n+n2B) = O (n2B)
- Where n= number of agents,
- B= number of Fake objects.

S-Random Implementation:

In this module the a lot of knowledge objects the agents request in total, the a lot of recipients on the average An object has and also the a lot of objects area unit shared among completely different agents, the harder it's to discover a guilty agent. During this rule, the agent receives solely the set of knowledge object which will tend to the agent

Data Distributor:

A knowledge distributor has given sensitive data to a group of purportedly sure agents (third parties). a number of the information is leaked and located in associate unauthorized place (e.g., on the online or somebody's laptop). The distributor should assess the chance that the leaked knowledge came from one or additional agents, as against having been severally gathered by different means that.

E-Mail Filtering:

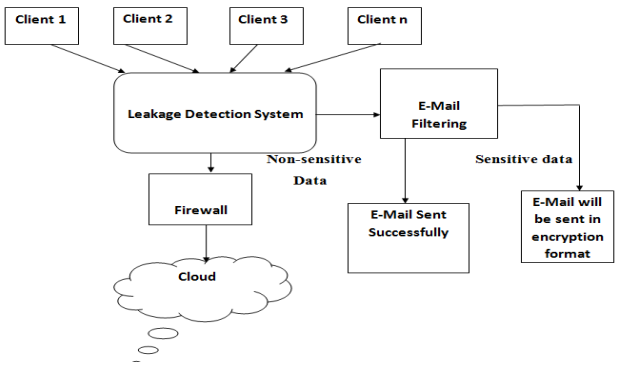
This module involves 6 steps.

1. Identify the data.
2. Remove stopping words such as this, is, a, etc.
3. Remove or change the synonyms.
4. Calculate the priority of the word depending upon the sensitivity of the data.
5. Compare data with predefine company datasets.
6. Filter the data if it has company's important data sets.

Instant Mail Alert:

The owner of the organization will get the mail alert of the data leaker (or) guilt agent.

5. SYSTEM ARCHITECTURE:



6. ALGORITHM DETAILS

S-Random:

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 1, Issue 10, October 2014)

In s-random, we have a tendency to introduce vector that shows the article sharing distribution. Specifically, shows the amount of agents United Nations agency receive object tk. algorithmic rule s-random allocates objects to agents in a very round-robin fashion. When the formatting of vectors d and a in lines one and a couple of of algorithmic rule four, the most loop in lines is dead whereas there are still knowledge objects to be allotted to agents. In every iteration of this loop, the algorithmic rule uses perform SELECT OBJECT () to search out a random object to portion to agent Uri. This loop iterates over all agents United Nations agency haven't received the amount of knowledge objects they need requested. The period of time of the algorithmic rule is and depends on the period of time nine of the article choice perform SELECT OBJECT().

Procedure:

Input: m1; . . .;mn, jTj . Assuming mi _ jTj
Output: R1; . . .;Rn
1: a 0jTj . a½k_: number of agents who have
received object tk
2: R1 ;; . . .;Rn ;
3: remaining Pn
i¼1 mi
4: while remaining > 0 do
5: for all i ¼ 1; . . . ; n : jRij <mi do
6: k SELECTOBJECTði;RiÞ . May also use
Additional parameters
7: Ri Ri [ftkg
8: a½k_ a½k_ þ 1
9: remaining remaining _ 1

E-Random:

In issues of sophistication EF, the distributor isn't allowed to feature pretend objects to the distributed information. So, the information allocation is totally outlined by the agents' data requests. Therefore, there's nothing to optimize. In EF issues, objective values square measure initialized by agents' information requests. Say, for instance, the distributor will add one pretend object to either R1 or R2 to extend the corresponding divisor of the summation term. Assume that the distributor creates a pretend object f and he provides it to agent R1. Formula one may be a general "driver" which will be employed by alternative ways, whereas formula two truly performs the random choice.

Procedure:

Input: R1; . . .;Rn, cond1; . . . ; condn, b1; . . . ; bn, B
Output: R1; . . .;Rn, F1; . . . ; Fn
1: R ; . Agents that can receive fake objects
2: for i ¼ 1; . . . ; n do
3: if bi > 0 then
4: R R [fig
5: Fi ;
6: while B > 0 do
i SELECTAGENTðR;R1; . . .;RnÞ
8: f CREATE FAKE OBJECTðRi; Fi; condiÞ
9: Ri Ri [ffg
10: Fi Fi [ffg
11: bi bi _ 1
12: if bi ¼ 0then
13: R RnfRig
14: B B _ 1

AES Algorithm

AES is based on a design principle known as a Substitution permutation network. It is fast in both software and hardware. Unlike its predecessor, DES, AES does not use a Feistel network.AES has a fixed block size of 128 bits and a key size of 128, 192, or 256 bits, whereas Rijndael can be specified with block and key sizes in any multiple of 32 bits, with a minimum of 128 bits. The block size has a maximum of 256 bits, but the key size has no theoretical maximum. AES operates on a 4×4 column-major order matrix of bytes, termed the state (versions of Rijndael with a larger block size have additional columns in the state). Most AES calculations are done in a special finite field. The AES cipher is specified as a number of repetitions of transformation rounds that convert the input plaintext into the final output of ciphertext. Each round consists of several processing steps, including one that depends on the encryption key. A set of reverse rounds are applied to transform ciphertext back into the original plaintext using the same encryption key.

High-level description of the algorithm

1) Key Expansion—round keys are derived from the cipher key using Rijndael's key schedule.
2) Initial Round:
AddRoundKey—each byte of the state is combined with the round key using bitwise xor
3) Rounds
a)SubBytes—a non-linear substitution step where each byte is replaced with another according to a lookup table.
b) ShiftRows—a transposition step where each row of the state is shifted cyclically a certain number of steps.
c) MixColumns—a mixing operation which operates on the columns of the state, combining the four bytes in each column.
D) AddRoundKey
4) Final Round (no MixColumns)
SubBytes
ShiftRows
AddRoundKey

MIX column

State1,1=GF(2,0x01) ⊕ GF(3,0x05) ⊕ GF(1,0x09)

Diagrams

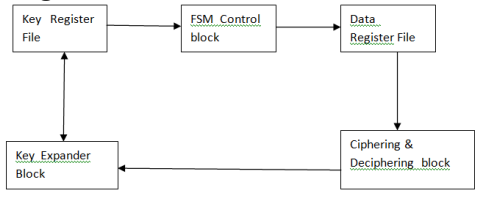
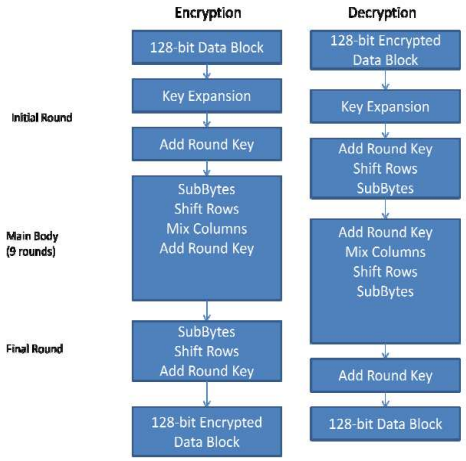


Fig: - Algorithm Flow Chart

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 1, Issue 10, October 2014)



CONCLUSION:

In this paper, we've projected AN approach that identifies that a part of intermediate information sets must be encrypted where as the remainder doesn't, so as to avoid wasting the privacy conserving value. A tree structure has been sculptured from the generation relationships of intermediate in formation sets to the investigate Privacy propagation among information sets. We have cultured downside the matter of saving privacy preserving value as a forced improvement problem that is addressed by rotten the privacy run constraints. A sensible heuristic algorithmic program has been designed consequently. Analysis results on real world information sets and bigger intensive information sets with numerous information and computation intensive applications on cloud, intermediate information set management is changing into a crucial analysis space. Privacy conserving for intermediate information sets is one amongst necessary however difficult analysis problems, and desires intensive investigation. With the contributions of this paper, we tend to are reaching to any investigate privacy aware economical programming of intermediate information sets in cloud by taking privacy conserving as a metric along side different metrics like s to rage and computation. Optimized balanced programming ways are expected to be developed toward over all extremely economical privacy ware data is set for programming.

REFERENCES:

[1]. M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A.Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M.Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53,no. 4, pp. 50-58, 2010.

[2]. R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, and I. Brandic,"Cloud Computing and Emerging It Platforms: Vision, Hype, and Reality for Delivering Computing as the Fifth Utility," Future Generation Computer Systems, vol. 25, no. 6, pp. 599-616, 2009.

[3]. L. Wang, J. Zhan, W. Shi, and Y. Liang, "In Cloud, Can Scientific Communities Benefit from the Economies of Scale?," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 2, pp. 296-303, Feb.2012.

[4]. H. Takabi, J.B.D. Joshi, and G. Ahn, "Security and Privacy Challenges in Cloud Computing Environments," IEEE Security & Privacy, vol. 8, no. 6, pp. 24-31, Nov./Dec. 2010.

[5]. D. Zissis and D. Lekkas, "Addressing Cloud Computing Security Issues," Future Generation Computer Systems, vol. 28, no. 3, pp. 583-592, 2011.

[6]. D. Yuan, Y. Yang, X. Liu, and J. Chen, "On-Demand Minimum Cost Benchmarking for Intermediate Data Set Storage in Scientific Cloud Workflow Systems," J. Parallel Distributed Computing, vol. 71, no. 2, pp. 316-332, 2011.

[7]. S.Y. Ko, I. Hoque, B. Cho, and I. Gupta, "Making Cloud Intermediate Data Fault-Tolerant," Proc. First ACM Symp. Cloud Computing (SoCC '10), pp. 181-192, 2010.

[8]. H. Lin and W. Tzeng, "A Secure Erasure Code-Based Cloud Storage System with Secure Data Forwarding," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 6, pp. 995-1003, June 2012.

[9]. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM '11, pp. 829-837, 2011.

[10]. M. Li, S. Yu, N. Cao, and W. Lou, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing," Proc. 31st Int'l Conf. Distributed Computing Systems (ICDCS '11), pp. 383-392, 2011.

[11]. C. Gentry, "Fully Homomorphic Encryption Using Ideal Lattices," Proc. 41st Ann. ACM Symp. Theory of Computing (STOC '09), pp. 169-178, 2009.

[12]. B.C.M. Fung, K. Wang, and P.S. Yu, "Anonymizing Classification Data for Privacy Preservation," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 5, pp. 711-725, May 2007.

[13]. B.C.M. Fung, K. Wang, R. Chen, and P.S. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," ACM Computing Survey, vol. 42, no. 4, pp. 1-53, 2010.

[14]. X. Zhang, C. Liu, J. Chen, and W. Dou, "An Upper-Bound Control Approach for Cost-Effective Privacy Protection of Intermediate Data Set Storage in Cloud," Proc. Ninth IEEE Int'l Conf. Dependable, Autonomic and Secure Computing (DASC '11), pp. 518-525, 2011.

[15]. I. Roy, S.T.V. Setty, A. Kilzer, V. Shmatikov, and E. Witchel, "Airavat: Security and Privacy for Mapreduce," Proc. Seventh USENIX Conf. Networked Systems Design and Implementation (NSDI '10), p. 20, 2010.

[16]. K.P.N. Puttaswamy, C. Kruegel, and B.Y. Zhao, "Silverline: Toward Data Confidentiality in Storage-Intensive Cloud Applications," Proc. Second ACM Symp. Cloud Computing (SoCC '11), 2011.

[17]. K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, "Sedic: Privacy-Aware Data Intensive Computing on Hybrid Clouds," Proc. 18th ACM Conf. Computer and Comm. Security (CCS '11), pp. 515-526, 2011.

[18]. V. Ciriani, S.D.C.D. Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, and P. Samarati, "Combining Fragmentation and Encryption to Protect Privacy in Data Storage," ACM Trans. Information and System Security, vol. 13, no. 3, pp. 1-33, 2010.

[19]. S.B. Davidson, S. Khanna, T. Milo, D. Panigrahi, and S. Roy, "Provenance Views for Module Privacy," Proc. 30th ACM SIGMOD-SIGACT-SIGARTSymp. Principles of Database Systems (PODS '11), pp. 175-186, 2011.

[20]. [20] S.B. Davidson, S. Khanna, S. Roy, J. Stoyanovich, V. Tannen, and Y. Chen, "On Provenance and Privacy," Proc. 14th Int'l Conf. Database Theory, pp. 3-10, 2011.