International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

Multi Keyword Based Cloud Data Encryption

M. Anusha & M.Venu Gopal

1. M. Tech Scholar, Department of IT, J.B. Institute of Engineering & Technology, Hyderabad, Telangana.

2. Assistant Professor, Department of IT, J.B. Institute of Engineering & Technology, Hyderabad, Telangana.

ABSTRACT

Due to The Increasing Popularity Of Cloud Computing, More And More Data Owners Are Motivated To Outsource Their Data To Cloud Servers For Great Convenience And Reduced Cost In Data Management. In any case, Sensitive Data Should Be Encrypted Before Outsourcing For Privacy Requirements, Which Obsoletes Data Utilization Like Keyword-Based Document Retrieval. In This Paper, We Present A Secure Multi-Keyword Ranked Search Scheme Over Encrypted Cloud Data, Which Simultaneously Supports Dynamic Update Operations Like Deletion And Insertion Of Documents. Specifically, The Vector Space Model And The Widely-Used TF_IDF Model Are Combined In The Index Construction And Query Generation. We Construct A Special Tree-Based Index Structure And Propose A "Greedy Depth-First Search" Algorithm To Provide Efficient Multi-Keyword Ranked Search. The Secure Knn Algorithm Is Utilized To Encrypt The Index And Query Vectors, And Meanwhile Ensure Accurate Relevance Score Calculation Between Encrypted Index And Query Vectors. In Order To Resist Statistical Attacks, Phantom Terms Are Added To The Index Vector For Blinding Search Results. As a result of The Use Of Our Special Tree-Based Index Structure, The Proposed Scheme Can Achieve Sub-Linear Search Time And Deal With The Deletion And Insertion Of Documents Flexibly. Expansive Experiments Are Conducted To Demonstrate The Efficiency Of The Proposed Scheme.

INTRODUCTION

Distributed computing Is The Use Of Computing Resources (Hardware And Software) That Are Delivered As A Service Over A Network (Typically The Internet). The Name Comes From The Common Use Of A Cloud-Shaped Symbol As An Abstraction For The Complex Infrastructure It Contains In System Diagrams. Disseminated figuring Entrusts Remote Services With A User's Data, Software And Computation. Distributed computing Consists Of Hardware And Software Resources Made Available On The Internet As Managed Third-Party Services. These Services Typically Provide Access To

Advanced Software Applications And High-End Networks Of Server Computers.

1.1 How Cloud Computing Works?

➢ The Goal Of Cloud Computing Is To Apply Traditional Supercomputing, Or High-Performance Computing Power, Normally Used By Military And Research Facilities, To Perform Tens Of Trillions Of Computations Per Second, In Consumer-Oriented Applications Such As Financial Portfolios, To Deliver Personalized Information, To Provide Data Storage Or To Power Large, Immersive Computer Games. The Cloud Computing Uses Networks Of Large Groups Of Servers Typically Running Low-Cost Consumer PC Technology With Specialized Connections To Spread Data-Processing Chores Across Them. This Shared IT Infrastructure Contains Large Pools Of Systems That Are Linked Together. Often, Virtualization Techniques Are Used To Maximize The Power Of Cloud Computing.

On-Demand Self-Service: A Consumer Can Unilaterally Provision Computing Capabilities, Such As Server Time And Network Storage, As Needed Automatically Without Requiring Human Interaction With Each Service's Provider.

Wide Network Access: Capabilities Are Available Over The Network And Accessed Through Standard Mechanisms That Promote Use By Heterogeneous Thin Or Thick Client Platforms (E.G., Mobile Phones, Laptops, And Pdas).

Resource Pooling: The Provider's Computing Resources Are Pooled To Serve Multiple Consumers Using A Multi-Tenant Model, With Different Physical And Virtual

Resources Dynamically Assigned And Reassigned According To Consumer Demand. There Is A Sense Of Location-Independence In That The Customer Generally Has No Control Or Knowledge Over The Exact Location Of The Provided Resources But May Be Able To Specify Location At A Higher Level Of Abstraction (E.G., Country, State, Or Data Center). Occurrences Of Resources Include Storage, Processing, Memory, Network Bandwidth, And Virtual Machines.

Quick Elasticity: Capabilities Can Be Rapidly And Elastically Provisioned, once in a while Automatically, To Quickly Scale Out And Rapidly Released To Quickly Scale In. To The Consumer, The Capabilities Available For Provisioning Often Appear To Be Unlimited And Can Be Purchased In Any Quantity At Any Time.

Measured Service: Cloud Systems Automatically Control And Optimize Resource Use By Leveraging A Metering Capability At Some Level Of Abstraction Appropriate To The Type Of Service (E.G., Storage, Processing, Bandwidth, And Active User Accounts). Asset Usage Can Be Managed, Controlled, And Reported Providing Transparency for both the Provider and Consumer of the Utilized

LITERATURE SURVEY

Security Challenges For The Public Cloud

Coursed figuring Represents Today's Most Exciting Computing Paradigm Shift In Information Technology. In any case, Security And Privacy Are Perceived As Primary Obstacles To Its Wide Adoption. Here, The Authors Outline Several Critical Security Challenges And Motivate Further Investigation Of Security Solutions For A Trustworthy Public Cloud Environment.

A Fully Homomorphic Encryption Scheme

We Propose The First Fully Homomorphic Encryption Scheme, Solving An Old Open Problem. Such A Scheme Allows One To Compute Arbitrary Functions Over Encrypted Data Without The Decryption Key—I.E., Given Encryptions E(M1), ..., E(Mt) Of M1, ..., M T, One Can Efficiently Compute A Compact Ciphertext That Encrypts F(M1, ..., M T) For Any Efficiently Computable Function

Absolutely Homomorphic Encryption Has Numerous Applications. For Example, It Enables Encrypted Search Engine Queries—I.E., A Search Engine Can Give You A Succinct Encrypted Answer To Your (Boolean) Query Without Even Knowing What Your Query Was. It Also Enables Searching On Encrypted Data; You Can Store Your Encrypted Data On A Remote Server, And Later Have The Server Retrieve Only Files That (When Decrypted) Satisfy Some Boolean Constraint, Even Though The Server Cannot Decrypt The Files On Its Own. All the more Broadly, It Improves The Efficiency Of Secure Multiparty Computation.

Public Key Encryption With Keyword Search

We Study The Problem Of Searching On Data That Is Encrypted Using A Public Key System. Consider User Bob Who Sends Email To User Alice Encrypted Under Alice's Public Key. An Email Gateway Wants To Test Whether The Email Contains The Keyword "Crushing" So That It Could Route The Email Accordingly. Alice, On The Other Hand Does Not Wish To Give The Gateway The Ability To Decrypt All Her Messages.

Practical Techniques For Searches On Encrypted Data

It Is Desirable To Store Data On Data Storage Servers Such As Mail Servers And File Servers In Encrypted Form To Reduce Security And Privacy Risks. Be that as it may, This Usually Implies That One Has To Sacrifice Functionality For Security. For Example, If A Client Wishes To Retrieve Only Documents Containing Certain Words, It Was Not Previously Known How To Let The Data Storage Server Perform The Search And Answer The Query, Without Loss Of Data Confidentiality. We Describe Our Cryptographic Schemes For The Problem Of Searching On Encrypted Data And Provide Proofs Of Security For The Resulting Crypto Systems. Our Techniques Have A Number Of Crucial Advantages.

Privacy Preserving Keyword Searches On Remote Encrypted Data

We Consider The Following Problem: A User U Wants To Store His Files In An Encrypted Form On A Remote File Server S. Later The User U Wants To Efficiently Retrieve Some Of The Encrypted Files Containing (Or Indexed By) Specific Keywords, Keeping The Keywords Themselves

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

Secret And Not Jeopardizing The Security Of The Remotely Stored Files.

SYSTEM ANALYSIS

EXISTING SYSTEM:

❖ A General Approach To Protect The Data Confidentiality Is To Encrypt The Data Before Outsourcing.

❖ Searchable Encryption Schemes Enable The Client To Store The Encrypted Data To The Cloud And Execute Keyword Search Over Ciphertext Domain. Up to this point, Abundant Works Have Been Proposed under Different Threat Models to Achieve Various Search Functionality, Such As Single Keyword Search, Similarity Search,

❖ Multi-Keyword Boolean Search, Ranked Search, Multi-Keyword Ranked Search, Etc. Among Them, Multi-Keyword Ranked Search Achieves More And More Attention For Its Practical Applicability.

❖ Recently, Some Dynamic Schemes Have Been Proposed To Support Inserting And Deleting Operations On Document Collection. These Are Significant Works As It Is Highly Possible That The Data Owners Need To Update Their Data On The Cloud Server.

DAMAGE OF EXISTING SYSTEM:

❖ Huge Cost In Terms Of Data Usability. For Example, The Existing Techniques On Keyword-Based Information Retrieval, Which Are Widely Used On The Plaintext Data, Cannot Be Directly Applied On The Encrypted Data. Downloading All The Data From The Cloud And Decrypt Locally Is Obviously Impractical.

❖ Existing System Methods Not Practical Due To Their High Computational Overhead For Both The Cloud Sever And User.

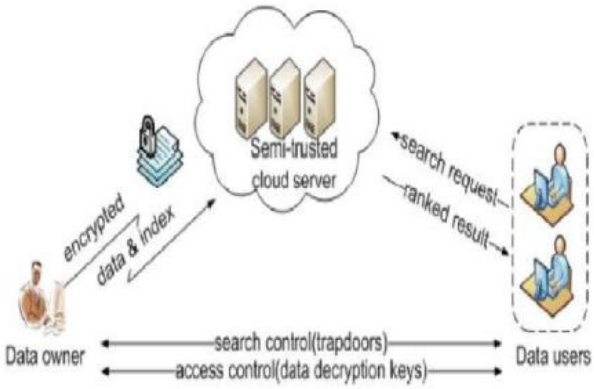
4.2 PROPOSED SYSTEM:

❖ This Paper Proposes A Secure Tree-Based Search Scheme Over The Encrypted Cloud Data, Which Supports Multi-Keyword Ranked Search And Dynamic Operation On The Document Collection. Specifically, The Vector Space Model And The Widely-Used "Term Frequency (TF) × Inverse Document Frequency (IDF)" Model Are Combined In The Index Construction And Query Generation To Provide Multi-Keyword Ranked Search. In Order To Obtain High Search Efficiency, We Construct A Tree-Based Index Structure And Propose A "Greedy Depth-First Search" Algorithm Based On This Index Tree.

❖ The Secure Knn Algorithm Is Utilized To Encrypt The Index And Query Vectors, And Meanwhile Ensure Accurate Relevance Score Calculation Between Encrypted Index And Query Vectors.

❖ To Resist Different Attacks In Different Threat Models, We Construct Two Secure Search Schemes: The Basic Dynamic Multi-Keyword Ranked Search (BDMRS) Scheme In The Known Ciphertext Model, And The Enhanced Dynamic Multi-Keyword Ranked Search (EDMRS) Scheme In The Known Background Model.

SYSTEM DESIGN



DATA FLOW DIAGRAM:

1. The DFD Is Also Called As Bubble Chart. It Is A Simple Graphical Formalism That Can Be Used To Represent A System In Terms Of Input Data To The System, Various Processing Carried Out On This Data, And The Output Data Is Generated By This System.

2. The Data Flow Diagram (DFD) Is One Of The Most Important Modeling Tools. It Is Used To Model The System Components. These Components Are The System Process, The Data Used By The Process, An External Entity That Interacts With The System And The Information Flows In The System.

3. DFD Shows How The Information Moves Through The System And How It Is Modified By A Series Of Transformations. It Is A Graphical Technique That Depicts Information Flow And The Transformations That Are Applied As Data Moves From Input To Output.

4. DFD Is Also Known As Bubble Chart. A DFD May Be Used To Represent A System At Any Level Of Abstraction. DFD May Be Partitioned Into Levels That Represent Increasing Information Flow And Functional Detail.

UML DIAGRAMS

UML Stands For Unified Modeling Language. UML Is A Standardized General-Purpose Modeling Language In The Field Of Object-Oriented Software Engineering. The Standard Is Managed, And Was Created By, The Object Management Group.

The Goal Is For UML To Become A Common Language For Creating Models Of Object Oriented Computer Software. In Its Current Form UML Is Comprised Of Two Major Components: A Meta-Model And A Notation. In The Future, Some Form Of Method Or Process May Also Be Added To; Or Associated With, UML.

The Unified Modeling Language Is A Standard Language For Specifying, Visualization, Constructing And Documenting The Artifacts Of Software System, As Well As For Business Modeling And Other Non-Software Systems.

The UML Represents A Collection Of Best Engineering Practices That Have Proven Successful In The Modeling Of Large And Complex Systems.

The UML Is A Very Important Part Of Developing Objects Oriented Software And The Software Development Process. The UML Uses Mostly Graphical Notations To Express The Design Of Software Projects.

USE CASE DIAGRAM:

A Use Case Diagram In The Unified Modeling Language (UML) Is A Type Of Behavioral Diagram Defined By And Created From A Use-Case Analysis. Its Purpose Is To Present A Graphical Overview Of The Functionality Provided By A System In Terms Of Actors, Their Goals (Represented As Use Cases), And Any Dependencies Between Those Use Cases. The Main Purpose Of A Use Case Diagram Is To Show What System Functions Are Performed For Which Actor. Parts Of The Actors In The System Can Be Depicted.

MODULES

This Module Helps The Owner To Register Those Details And Also Include Login Details. This Module Helps The Owner To Upload His File With Encryption Using RSA Algorithm. This Ensures The Files To Be Protected From Unauthorized User. Data Owner Has A Collection Of Documents F ={F1; F2; :::; Fn} That He Wants To Outsource To The Cloud Server In Encrypted Form While Still Keeping The Capability To Search On Them For Effective Utilization.

In Our Scheme, the Data Owner Firstly Builds A Secure Searchable Tree Index I from Document Collection F, And Then Generates An Encrypted Document Collection C For F. From there on, The Data Owner Outsources the Encrypted Collection C and the Secure Index I to the Cloud Server, And Securely Distributes the Key Information of Trapdoor Generation and Document Decryption to the Authorized Data Users. In addition, The Data Owner Is Responsible for the Update Operation of His Documents Stored in the Cloud Server. While Updating, The Data Owner Generates The Update Information Locally And Sends It To The Server.

Data User Module

This Module Includes The User Registration Login Details. This Module Is Used To Help The Client To Search The File Using The Multiple Key Words Concept And Get The Accurate Result List Based On The User Query. The User Is Going To Select The Required File And Register The User Details And Get Activation Code.

In Mail Email Before Enter The Activation Code. After User Can Download The Zip File And Extract That File. Data Users Are Authorized Ones To Access The Documents Of Data Owner. With T Query Keywords, The Authorized User Can Generate A Trapdoor TD According

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

To Search Control Mechanisms To Fetch K Encrypted Documents From Cloud Server. By then, The Data User Can Decrypt The Documents With The Shared Secret Key.

Rank Search Module

These Modules Ensure The User To Search The Files That Are Searched Frequently Using Rank Search. This Module Allows The User To Download The File Using His Secret Key To Decrypt The Downloaded Data. This Module Allows The Owner To View The Uploaded Files And Downloaded Files. The Proposed Scheme Is Designed To Provide Not Only Multi-Keyword Query And Accurate Result Ranking, But Also Dynamic Update On Document Collections. The Scheme Is Designed To Prevent The Cloud Server From Learning Additional Information About The Document Collection, The Index Tree, And The Query

SYSTEM TESTING

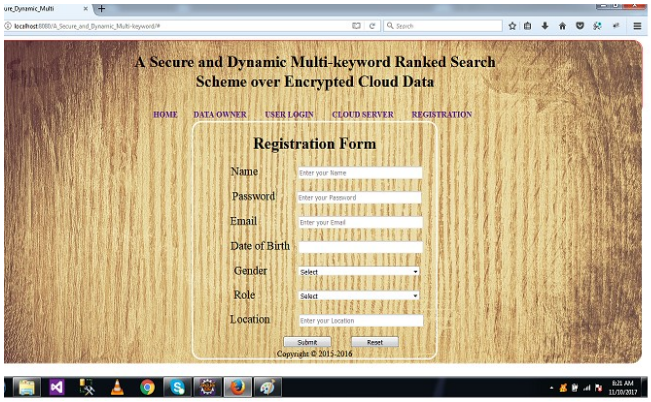
The Purpose Of Testing Is To Discover Errors. Testing Is The Process Of Trying To Discover Every Conceivable Fault Or Weakness In A Work Product. It Provides A Way To Check The Functionality Of Components, Sub Assemblies, Assemblies And/or A Finished Product It Is The Process Of Exercising Software With The Intent Of Ensuring That The

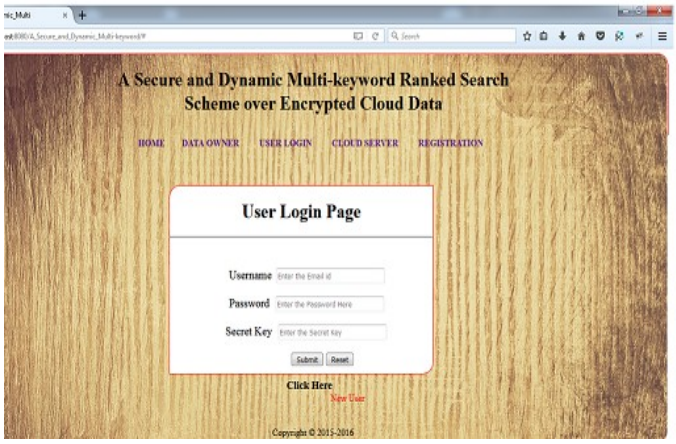
Programming System Meets Its Requirements And User Expectations And Does Not Fail In An Unacceptable Manner. There Are Various Types Of Test. Each Test Type Addresses A Specific Testing Requirement.

Sorts OF TESTS

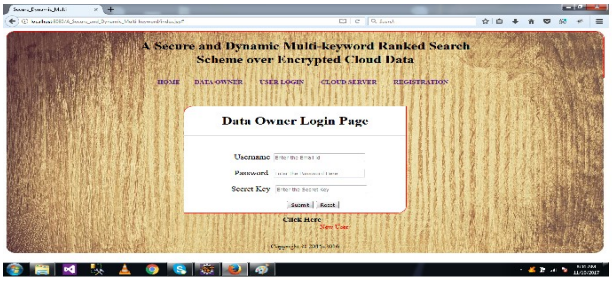
Unit Testing
Mix Testing
Rational Test
Framework Test
White Box Testing
Exposure Testing

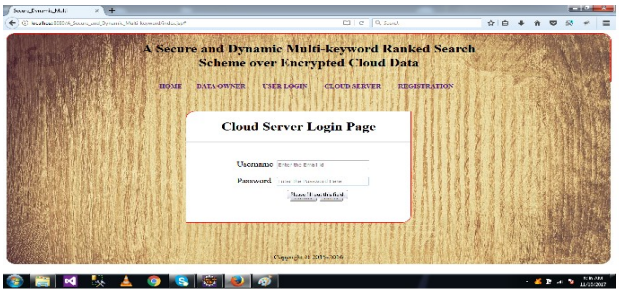
RESULT SCREEN SHOTS





International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017





CONCLUSION

In This Paper, A Secure, Efficient And Dynamic Search Scheme Is Proposed, Which Supports Not Only The Accurate Multi-Keyword Ranked Search But Also The Dynamic Deletion And Insertion Of Documents. We Construct A Special Keyword Balanced Binary Tree As The Index, And Propose An "Eager Depth-First Search" Algorithm To Obtain Better Efficiency Than Linear Search. In Addition, The Parallel Search Process Can Be Carried Out To Further Reduce The Time Cost. The Security Of The Scheme Is Protected Against Two Threat Models By Using The Secure Knn Algorithm. Trial Results Demonstrate The Efficiency Of Our Proposed Scheme.

There Are Still Many Challenge Problems In Symmetric SE Schemes. In The Proposed Scheme, The Data Owner Is Responsible For Generating Updating Information And Sending Them To The Cloud Server. Consequently, The Data Owner Needs To Store The Unencrypted Index Tree And The Information That Are Necessary To Recalculate The IDF Values. Such An Active Data Owner May Not Be Very Suitable For The Cloud Computing Model. It Could Be A Meaningful But Difficult Future Work To Design A Dynamic Searchable Encryption Scheme Whose Updating Operation Can Be Completed By Cloud Server Only, Meanwhile Reserving The Ability To Support Multi-Keyword Ranked Search.

In Addition, As The Most Of Works About Searchable Encryption, Our Scheme Mainly Considers The Challenge From The Cloud Server. Everything considered, There Are Many Secure Challenges In A Multi-User Scheme. At to begin with, All The Users Usually Keep The Same Secure Key For Trapdoor Generation In A Symmetric SE Scheme. In This Case, The Revocation Of The User Is Big Challenge.

On the off chance that It Is Needed To Revoke A User In This Scheme, We Need To Rebuild The Index And Distribute The New Secure Keys To All The Authorized Users. Moreover, Symmetric SE Schemes Usually Assume That All The Data Users Are Trustworthy. It Is Not Practical And A Dishonest Data User Will Lead To Many Secure Problems. For Example, A Dishonest Data User May Search The Documents And Distribute The Decrypted Documents To The Unauthorized Ones. Altogether More, A Dishonest Data User May Distribute His/her Secure Keys To The Unauthorized Ones. In The Future Works, We Will Try To Improve The SESchemeToHandleTheseChallengeProblems.

REFERENCES

[1] K. Ren, C.Wang, Q.Wang Et Al., “Security Challenges For The Public Cloud,” IEEE Internet Computing, Vol. 16, No. 1, Pp. 69–73, 2012.
[2] S. Kamara And K. Lauter, “Cryptographic Cloud Storage,” In Financial Cryptography And Data Security. Springer, 2010, Pp. 136–149.
[3] C. Gentry, “A Fully Homomorphic Encryption Scheme,” Ph.D. Dissertation, Stanford University, 2009.
[4] O. Goldreich And R. Ostrovsky, “Software Protection And Simulation On Oblivious Rams,” Journal Of The ACM (JACM), Vol. 43, No. 3, Pp. 431–473, 1996.

[5] D. Boneh, G. Di Crescenzo, R. Ostrovsky, And G. Persiano, "Public Key Encryption With Keyword Search," In Advances In Cryptology-Eurocrypt 2004. Springer, 2004, Pp. 506–522.
[6] D. Boneh, E. Kushilevitz, R. Ostrovsky, And W. E. Skeith III, "Public Key Encryption That Allows Pir Queries," In Advances In Cryptology-CRYPTO 2007. Springer, 2007, Pp. 50–67.
[7] D. X. Song, D. Wagner, And A. Perrig, "Practical Techniques For Searches On Encrypted Data," In Security And Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium On. IEEE, 2000, Pp. 44–55.
[8] E.-J. Goh Et Al., "Secure Indexes." IACR Cryptology Eprint Archive, Vol. 2003, P. 216, 2003.