PRIVACY-ENABLED MULTI-KEYWORD SEARCH OVER ENCRYPTED CLOUD DATA

Mekuria Gemechu

A thesis submitted to the Department of Computer Science and Engineering, School of Electrical Engineering and Computing in partial fulfillment of master's degree in Computer Science and Engineering

Office of Graduate Studies
Adama Science and Technology University
Adama, Oromia, Ethiopia

Adama
April 17, 2019

PRIVACY-ENABLED MULTI KEYWORD SEARCH OVER ENCRYPTED CLOUD DATA

Mekuria Gemechu Tufa

mekuriaossie1@gmail.com

Advisor: Dr. Ravindra Babu

A thesis submitted to the Department of Computer Science and Engineering, School of Electrical Engineering and Computing in partial fulfillment of master's degree in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

Adama, Oromia, Ethiopia

Adama

April 17, 2019

APPROVAL OF THE BOARD OF EXAMINERS

We, the undersigned, members of the Board of Examiners of the final open defense by Mekuria Gemechu have read and evaluated his thesis entitled "**Privacy-Enabled Multi-keyword Search Over Encrypted Cloud Data**" and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirement of the degree of Masters in Computer Science and Engineering (CSE).

Ravindra Babu (PHD) (Advisor)	Signature	Date
(Chairperson)	Signature	Date
(Internal Examiner)	Signature	Date
(External Examiner)	Signature	Date

ADVISOR'S APPROVAL SHEET

To: Computer Science and Engineering program

Subject: Thesis submission

This is to certify that the thesis entitled "**Privacy-Enabled Multi-keyword Search Over Encrypted Cloud Data**" submitted in partial fulfillment of the requirements for the degree of Masters in Computer Science and Engineering, the Graduate program of the Computer Science and Engineering Program, and has been carried out by Mekuria Gemechu Id. No GSR/0059/09, under my supervision. Therefore, I recommend that the student has fulfilled the requirements and hence hereby he can submit the thesis to the department.

Ravindra Babu (Ph.D.)
(Advisor) _____ _____
Signature Date

DECLARATION

I hereby declare that this MSc thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this thesis have been duly acknowledged.

Mekuria Gemechu Tufa

_____

(Signature)

This MSc thesis has been submitted for examination with my approval as a university advisor.

Ravindra Babu (Ph.D.)
(Advisor) _____

(Signature)

Submission Date: _____

ACKNOWLEDGMENT

In a position at the front, my thanks go to my GOD almighty who provide me everything that helps me to accomplish this thesis work as it was expected. If it was not for the GOD almighty help I could not have been here.

During the time of this work, I have got a polite and good teacher whom I can say also a good friend who helped me much in his guidance, my advisor **Dr. Ravindra Babu**. He has been always available whenever I need him providing me his advice, correction, direction, and even the resources needed to accomplish this work. I was so free and full of knowledge under his guidance. This helps me do my work freely without any tension and fear. To be honest I couldn't accomplish this work if it was not for his support too.

I am also happy to send my thanks for the help of cloud computing research SIG lab members, **Mr. Desta (MSc), Mr. Endale (MSc), and Mr. Anteneh (MSc)** for passing their golden time to give me assistance, aid, guidance and support wherever and whenever I asked them to put me on the truck to my work.

My acknowledgment would not be complete if I forget to express my thanks to my family who gave me every hope that keeps my strength. The goodness of my uncle **Mr. Ossie Tufa &** his wife **Mrs. Tayiba Mohammed** and the love and care of my sister **Buzunesh Tufa,** and her husband **Tariku Taressa**, and of course my grandparents, **Tufa kabeto** and **Bedatu Harsiso** who helped me in making me active and things not to get bored, equipped me with too much hopes, strength and future to be where am I now. I send my thanks to them for making me see my future which full of success and happiness.

GOD bless you **Megersa and Muberak** for every good thing you have done to me and you were always with me even in bad condition too. I want to thanks those peoples who have supported me in my good and bad time directly or indirectly.

ABSTRACT

As cloud data storage gets big and big, many users who own data that are too big in size are interested to move their data to a remote cloud. For the purpose of keeping users' sensitive data secure, these sensitive users' data must be encrypted in an offline stage. There exist many searchable encryption algorithms to prove data existence. In whatever way, the existing search algorithms do give too little or no consideration whether data users' queries were efficient or not in a system that supports many data owners. The methods developed in this work is a multiple keyword searches that supports search results in a ranked manner which is based on a tree structure. As the data that users are interested in outsourcing to the remote cloud may be of a big size, term frequency and inverted document frequency which is the most appropriate method for ranking results of users' queries according to certain similarity criteria are used. To make the cloud server accomplish a secure data retrieval without knowing any users' private data, a novel search algorithm that retains privacy and also which is lay on bi-linear mapping is developed. To attain a timely search, for every data owner, an index that is developed in a tree structure with additive order and privacy retaining method family is established. After then by using a Merkle tree algorithm to look for the files of interest cloud server combine these indexes effectively. Lastly, a test that is done on the security of the proposed system makes sure that the developed work is secure, and the performance test makes it clear that it is efficient.

Keywords: Multi-keyword ranked search, multiple data owners, security, cloud storage, privacy-enabled, cloud security.

CONTENTS PAGES

LIST OF FIGURES

FIGURES **PAGES**

LIST OF TABLES

TABLES **PAGES**

LIST OF ABBREVIATIONS

.NET	Dot Network
AES	Advanced Encryption Standard
API	Application Programming Interface
ASCI	American Standard Code for Information Interchange
CPU	Central Processing Unit
CSPs	Cloud Service Providers
DES	Data Encryption Standard
DFS	Depth First Search
ERD	Entity Relationship Diagram
GDFS	Greedy-Depth First Search
GHz	Giga-Hertz
HTML	Hypertext Markup language
IaaS	Infrastructure as Service
IBM	International Business Machines
IDE	Integrated Development Environment
IDF	Inverse Document Frequency
IEEE	Institute of Electrical and Electronics Engineers
IR	Information Retrieval
IS	Information System
IT	Information Technology
KB	Kilo-Byte
kNN	K-Nearest Neighbor
MRSE	Multi-keyword Search over Encrypted Cloud Data
NASA	National Aeronautics and Space Administration
NIST	National Institute of Standards and Technology

OS	Operating System
PaaS	Platform as a Service
PC	Personnel Computer
PEKS	Public Key Encryption Escorted By Keyword Searching
PEMKSE	Privacy-Enabled Multi-keyword Search over Encrypted Cloud Data
PEMS	Privacy-Enabled Multi-Keyword Search system over Encrypted Cloud Data
PIR	Private information retrieval
RAM	Random Access Memory
RSA	Rivest, Shamir, and Adelman
SaaS	Software as a Service
S-box	Rijndael Substitution Box
SLA	System level Agreement
SOA	Service-Oriented Architecture
SQL	Structural Query Language
SSL	Secure Socket Layer
TF	Term Frequency
WS	Web Services
XML	Extendable Markup Language
XOR	Exclusive OR

LIST OF NOTATIONS

C	The encrypted file collection of O_i, denoted as a set of d files $C = (C_1, C_2, \ldots, C_d)$.
F	The plaintext file collection of O_i, denoted as a set of d files $F = (F_1, F_2 \ldots, F_d)$.
I	The collection of data owners' index, denoted as a set of m indexes $I = (I_1, I_2, \ldots, I_m)$.
O	The collection of data owners denoted as a set of m data owners $O = (O_1, O_2, \ldots, O_m)$
U	Data users collection denoted as a set of n data users $U = (U_1, U_2, \ldots, U_n)$.
W	The collection of keywords denoted as a set of n keywords $W = (w_1, w_2, \ldots, w_n)$

CHAPTER ONE

1. INTRODUCTION

Cloud is a distributed resource that is effective to a very high degree because it is not only utilized by a big quantity of customers but also can be flexibly reached up on the needs' of the customers. It is given a name cloud as a result of that size is flexible depending on the users' need, no real boundary, and various location like a true cloud in nature, to whatever degree, it proves existence in the true world. An ordinary set of hardware, software, and services cannot be simply a real cloud. But, the cloud is just a combination and integration of information technologies that are too big in its size. As the new Innovation or new technologies continue emerging and become part of the group the size of the cloud enlarge increasingly. According to the definition of the national institute of science and technologies of the united state department of commerce definition, cloud computing is an example for proving ambiguous possible and easy, enables a distributed pool of settable computing at the time of demand using network access. Cloud computing is a separable part of the internet that makes easy for enterprises, different organizations, and individual users who own enormous amount of data to store their data to cloud even being in different physical location of the server to meet the aim of on-needed to quality application and services such as online file storing, networking for social sites, webmail and many others from a distributed pool of settable computing resources [1].

To summarize it, cloud computing is the composition of old computing models and interconnection technologies such as decentralized computing, side by side computing, service computing, technologies for storing network, idealization, balancing work, anytime availability and etc.

Too many advantages that cloud computing provides creates a great influence on its customers such as firms, private organizations, and individual customers to load their enormous amount of searchable file that they own to servers which are not in the same physical location as the users'. The first most advantages of the cloud is that it lets users move their data to the server which are even not in the same physical sites by using their end devices that can be connected to the internet for getting internet services rather than using their own devices for storing and even managing their enormous amount of data. Removing the expenses of buying resources and setting their configuration.

As the advantages that the cloud computing provides for its customers show advancements from day to day, the interests of customers for moving their data show progress in size. Mostly it's more dynamicity and reducing of expenses are the topmost advantages that stole the attention of its customers for moving the large amount of data they own to the server that not fully trusted but partially trusted. The integration of all these advantages and the newly developing services and technologies are the most influential reason for cloud users to move their data which may also be sensitive data to the remotely found cloud storage.

The data that the cloud users outsource to the cloud storage may contain files that are too sensitive and at the same time, the cloud server to which they are moving their sensitive data are not fully trusted. Because of that servers are not fully trusted it is sure that the users' will raise a security or privacy question for their data. So the most standard way of answering these users' questions are hiding their sensitive data.

The most selected methods for hiding sensitive data is through encryption. So privacy of the users' data is kept by encrypting these data before loading [2] to cloud in order to realize security and eliminate access that is not granted, and this proves confidentiality through the cloud.

Using encryption for the privacy purpose makes the way of accessing these data different from the way of accessing plaintext which means that the way used for reaching plaintext is not applicable for the encrypted cloud data. In order to realize the advantages of the cloud and its services an efficient way of data accessing these encrypted data have to be developed. In fact, different kinds of search algorithms over the ciphertext are developed [3]–[6]. During the past time, almost all of the developed algorithms are applicable for the precise query matching. They are good at picking the data from the cloud storage depending on the user-submitted queries. In the real world, however, it is more obvious to do retrieval depending on the users' submitted query similarity with a predefined feature in place of the presence of it.

To make data accessing for a document that is too big in size timely cloud storage must operate on returning results of users' query in a ranked manner rather than returning simply unordered results to the users. This helps the user in finding their files of interest rather than working on ranking the results in the whole matches [7]. Returning back the users' query results that are of only the top related results which in return eliminates or reduces the unwanted network flow is

the most attractive advantages of the rank search is to keep data privacy, this kind of ordered search operation, however, should not make them vulnerable to an authorized access or cloud server. At the time of increasing user searching ability to increase users' search result correctness, it is also mandatory for ordered search to include searches with many keywords. One keyword search most of the time produced very large amounts of results. As a common activity showed by latest day's web finding engines (e.g., searching Google), data users may need to provide many keywords rather than only one as the sign of their search demand to get the most related data.

Additional, each and every search term in the users' submitted search query is able to guide steeper and steeper bottom the search response ahead. Content checking [8], i.e., a lot of expected matches, is a timely closeness measure from such many search term semantics to refine the final similarity and utilized vastly with unencrypted data retrieval (IR) community. But, searching the cipher files which were stored in cloud storage has been no an easy task because of the security bottlenecks.

As shown in the literature review, accessible encryption [9-13] is the best method that accepts a cipher data like documents and lets users to privately reach the data which are of their demand with one search query. However, the access methods which supports only one search term in its query are unable to support searches that needs many key terms in their search terms. In this type of data access methods, it is a dream to get the expected results from that much data size in a cloud. Searching over to big size data with a search request that contains only one key term is not efficient in giving back what is expected. Even though some nearest works have been done to include many keyword access [14-17] as to try to increase making access method dynamic, they are up to now not meet the desired results for making results ordered according to specified criteria.

They do not tolerate many data owners and their ciphering algorithm are not up to date with new technologies. A way of developing a secure and timely data reaching that tolerated many keywords in the users' query that also support many data owner at the same time keeping sensitive data private is still a challenging unclosed difficulty. In this observe, the issue of search the usage of many seek terms on cipher cloud statistics (MRSE) for lots proprietors parallel with maintaining their data at ease is clarified and solved. From among one-of-a-kind many keyword

semantics, the well-timed similarity calculating of coordinate evaluating, i.e., plenty of closest outcomes, to retrieve effects which are the most closed to the favored outcomes. Specifically, inner document similarity [4], i.e., the quantity of that documents to the gaining access to query.

The introduced fee of this paper is written as follows: firstly, the project of many keywords ordered access on cipher cloud information is chosen, and the number of actual security wishes for consisting of cloud file accomplishing device are evolved. Secondly, formal which means for the security and what is wanted for the privacy of the hunt using the keyword on cipher statistics is provided. Thirdly, MRSE for plenty document proprietor schemes which are based totally on closeness degree of coordinating evaluating at the same time as pleasant extraordinary desires for privacy is proposed. Fourthly, the superior Encryption trendy method is used and applied to MRSE. Finally, the proposed machine is checked to be in time in retrieving documents and relaxed as proved through walking simulation suggests.

1.1. Motivation

Another tremendously important access characteristic is recorded exchange that's sending and receiving information in between each other. corporations such health facility or clinical keep a record in their consumer’s records which need to be accessed via the clients without problems to his/her precise documents from that a great deal of facts size which will get his top documents. Within the same manner, any worker in an employer should have the capability to look for his facts files loaded by different employees as well as the organizations.

The work that has advanced in a near time [17] used a seek term on cipher cloud statistics which is in the form of tree shape, addresses the numerous keyword get entry to hassle in a single facts proprietor architecture. However, their proposed scheme isn't green and relevant only for one owner of the information, which means that their paintings do now not used in case of many information proprietors.

1.2. Statement of the problem

1.2.1. System model

Systematic architecture for this proposed work is a group of 3 modules. These are statistics proprietor, statistics consumer, and cloud server. in this version, statistics proprietor is an entity which has facts which are too massive in size and which can be of more in numbers F that he/she

needs to transport it to the remotely observed cloud as a cause of sharing advantages of cloud computing. Then the owner earlier than being connected to the internet generates an index from his files after which via creating a random key he/she encrypts files to be loaded to cloud garage using his personal key to shape a ciphered form of his document C. finally, each proprietor moves both the cipher record and their corresponding indexes to the cloud garage for later use by way of himself or different authorized peoples.

Then the furnish between records proprietor and user are needed to permit other users to get admission to one's documents through sending the formulated seek phrases that contain many seek phrases in his/her queries for documents of call for to the cloud server. After accepting a formulated search terms for documents in their call for from customers, the cloud server seems for the similar and related documents of customers' hobby by checking through saved indexes. Finally, the cloud server brought again outcomes in an ordered way to the customers thru web browser they're the using.

1.2.2. Threat model

Shielding the cloud server from understanding any statistics about the real information is the goal of the proposed system. Right here, the statistics user and information owner are assumed to be depended on at the same time as, a cloud server is semi-depended on, which is similar with the structure in [10], [18], [19]. Because of this, the cloud server isn't interested in actively modifying the go with the flow of the message or diverting another kind of offerings from their everyday manner. The problem of secure ranked key-word seek talked in this work is as follows: the consumer seek result ought to be returned in accordance with a few ranked similarity criteria which is the variety of search key-word lifestyles in the report of customers' interest, to growth report retrieval accuracy for users without preceding understanding on the encrypted document collection C. something, cloud server have to examine not anything or little approximately the similarity standards as leak important touchy statistics against keyword privateers. to be able to lessen the bandwidth, the person may ship an option cost k at the side of the quest keywords (Tw) and the cloud server best returns back the pinnacle-ok most that as a whole lot as in all likelihood applicable to the consumer's search keyword w.

In a fashionable, many institutions and organizations load their greater important statistics within the cloud to maintain their statistics from infection and hacking. The benefit of new computing is

it looks deeply for cloud clients. Rank search complements framework ease of use by using normal coordinating statistics in a ranked manner with respect to sure significance criteria (e.g. – keyword frequency). As straightforwardly outsourcing importance scores will trickle an extremely good deal of sensitive records in opposition to the key-word protection, to resolve this hassle asymmetric encryption with the ranking effect of question records with a view to supplying just anticipated information is proposed.

1.2.3. Design goals

The machine design of privateers Enabled Multi-key-word to seek over Encrypted facts in cloud computing should gain subsequent dreams:

- ***Multi-keyword Ranked Search for Multiple Data Owners****:* the proposed scheme allows not most effective multi-keyword search over encrypted cloud facts however additionally enables the cloud server to go back the pinnacle-okay files in a ranked manner based at the frequency of the keywords within the record of interest.
- ***Retrieval accuracy****:* the similarity between the consumer question and the files in returned results, and the relevance of documents within the back outcomes.
- ***Data privacy***: which means documents are modified into cipher format earlier than outsourcing to the untrusted. That is achieved by encrypting the files to be outsourced the usage of symmetric cryptography (AES). This scheme is likewise useful in protecting facts from hackers even though the physical statistics is cracked.
- ***Query privacy***: the primary privateer's aim is to hold the cloud server towards learning any important facts approximately the encrypted documents and queries submitted by the consumer.

1.3. Research questions

In the long run, this work will answer the following studies query (i.e., through jogging the advanced prototype at the simulation).

- How multi key-word seek reduces the file retrieval time from cloud storage through querying the database the usage of multi-key-word as a search key?
- Does the multi-keyword seek to simplify the search over cloud facts storage by way of displaying results of a query so as of relevance to the hunt key phrases?

- Does multi-keyword seek improves file search retrieval time over encrypted cloud storage and additionally returns results for key phrases that would have intended similar to our entered keyword?

1.4. Objectives

This observe has its very own particular and popular targets that will be fulfilled at the end of its crowning glory. Both precise and well-known objectives are listed and defined as follows in their subsection:

1.4.1. General objectives

Enabling cloud customers to get admission to their record of the hobby from among encrypted cloud information whilst stopping cloud server from understanding something vital approximately the encrypted document and person queries (i.e., enabling privacy) in multi-information owner version is the general objective of this take a look at.

1.4.2. Specific objectives

Precise targets of this examination are indexed under as follows:

- To encrypt and decrypt statistics files in a manner that forestalls the cloud server from gaining knowledge of vital information about the statistics documents at the same time as consuming less time and price.
- To get the maximum comparable outcomes that as lots as feasible suits the person search queries by means of using multi-key-word search machine while preserving the privacy of encrypted documents and person submitted seek queries.
- To show the maximum applicable user seek consequences in a ranked way in line with the frequency of the consumer seek keys in his/her file of pastimes.
- To display only the pinnacle-ok consumer search consequences because it helps in lowering the bandwidth.

1.5. The significance of the study

As cloud computing technology retain rising, records proprietors are encouraged to shift their complex information control systems from nearby websites to the economic public cloud for remarkable flexibility and financial savings. However, for retaining records privateers, touchy

records, such as which include private snapshots, emails, and many others. Should be encrypted earlier than outsourcing, which in flip emerge as a challenge for classic data utilization that is based totally on plaintext key-word seek. For this reason, developing a search service, which shall us facts consumer to easily seek over encrypted cloud data is of paramount significance. because the variety of cloud users and documents within the cloud turns into so big in quantity, allowing a couple of key phrases within the search query and returning the outcomes in order in their similarity to these keywords are important services that give no time. a lot of these together trigger the want for studying multi key-word search over encrypted cloud records whilst protective the privateers of the sensitive information. The beneficiaries of these paintings are users who registered as felony participants of that system which is a private cloud.

1.6. The scope of the study

This study focuses on growing multi-key-word ranked seek over encrypted cloud statistics. From symmetric cryptography, advance Encryption preferred (AES) is used to shield statistics documents privacy thru encryption. It additionally makes use of keyword frequency (i.e., the range of key phrases exists within the asked files) for displaying search results in a ranked way. With the quest question, the person can ship k non-compulsory value to reduce the bandwidth through returning best pinnacle-k most relevant consequences to the person.

1.7. Limitation of the study

As this take a look at having its personal additional offerings that it could perform it additionally has some features that it cannot assist, that's the issue of the observed. those obstacles are the rate at which documents are retrieved relies upon at the internet speed, the facts owner responds to consumer requests only when he is online, the report retrieval time also depends on the important thing length of AES set of rules, sending of the decryption key to consumer isn't always supported thru medium like email, the gadget handiest makes use of the garage component from cloud, key control is not performed in the system, and the system does now not guide uploading and downloading of documents like video.

1.8. Challenge

Other than the unmarried facts owner scheme, developing a cozy and efficient seek gadget for more than one information proprietors grow to be a difficult mission. To implement privacy-

permit and green searches, an index shape that's based totally on tree structure is built for every information proprietor’s encrypted records. For a unique query circumstance, facts users ought to create a search question for each records owner, and the cloud needs to also seek thru each index. This is not efficient, as a purpose for the linear courting of the wide variety of search queries and information proprietors. An easy way to run off this issue is to permit every records owner to use the equal key to encrypt their facts documents. Despite, any one of the proprietors compromised would possibly take to gadget failure.

1.9. Approach

In this paper, a couple of records proprietors are taken into consideration, via which each information proprietor creates a tree-based index for his/her documents and encrypts these files their corresponding key. To enforce a privateers-enabled and efficient seek device collectively, a privateers-Enabled Multi-keyword seeks gadget over Encrypted Cloud facts is proposed (PEMKSE). In this system, the cloud server is granted to effectively and securely carry out the ranked multi keyword search while retaining the proprietors’ sensitive records privateers in addition to the records documents and the queries. To show the search effects in a ranked way, TF-IDF scheme to version similarity score of records files are used. Ultimately, the proposed device is showed to be efficient and relaxed thru theoretical analysis and experiments with pattern information on simulation.

1.10. Roadmap

The coming section of this paper is arranged as follows: chapter two presents and briefly discusses the literature review (i.e. study background) and topmost related works. Chapter three presents the methodology followed during the study process. Chapter four presents the proposed work of the study. Chapter five presents the results of the study. Chapter six describes and presents future work and conclusion.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Cloud computing

In the past 10 years, the internet has been growing at a high rate. The cost of storing data or files, the power that the computer and hardware utilized is also increasing. The storage space of the data center cannot satisfy the needs of the customer and the system and service of original internet unable to solve the problem raised as a result of high demand for storage, so the new solution is needed. Simultaneously, large companies have to deal with data source fully to support their business [20].

Cloud computing is the delivery of the computing services which ranges from applications to storage and processing as they are needed by the customer using the internet and on a pay-as-you-go basis [21]. In cloud computing, there is no need for companies to buy or have their own computing infrastructure and data centers as they can get access to anything from application to storage from a cloud service provider through rent.

What is the need for cloud computing? It is to use the vacant resources of a computer, increasing economic efficiency by improving access rate, and reduce equipment energy consumption rate.

2.1.1. Cloud computing principle

It is not that much easy to define cloud as we want. Computing is not a real network of computing resources. It provides computing equipment's on the internet for users via the network of networks. Inter-combined cloud computing is a fully changeable computing system. It gives a must application software environment [22]. By saying this we mean that it can be settable, gives or re-provide computing equipment variable and follow the utilization of these resources always. Generally speaking, cloud computing contains a decentralized formation establishment, and watch over the decentralized system, to realize its purpose of making resource usage efficient [23].

Cloud computing gathers all the computing resources and watches over them automatically with the help of software. At the time data analysis, it combines the old data and current data to make

the gathered information more accurate and provide more intelligent services for users and companies [24]. No need for customers to care about how to buy servers, software, solutions and so on. Users can purchase the computing resource via the internet depending on their needs.

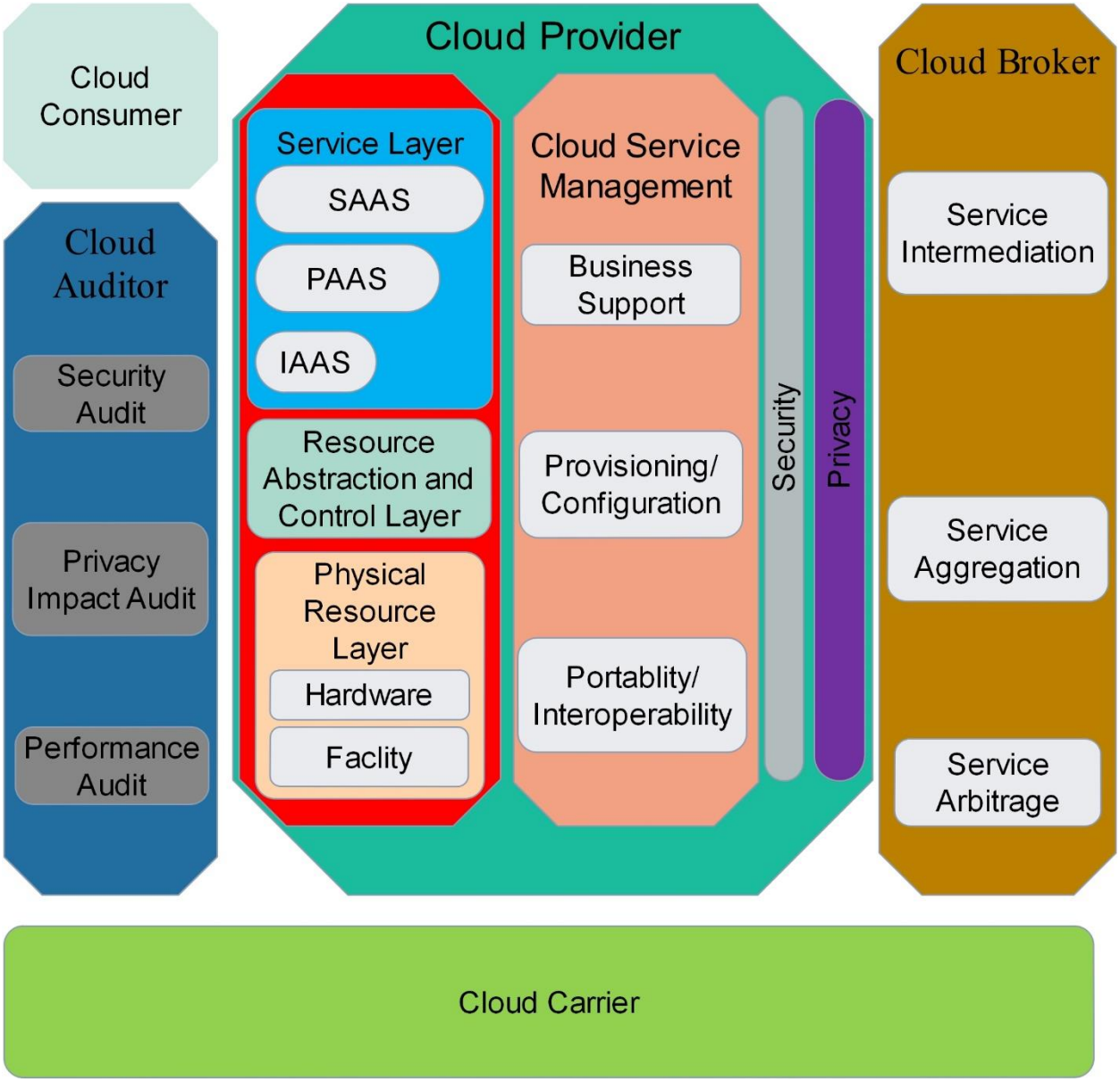


Figure 1. Cloud computing reference architecture [28]

Cloud computing does not rely upon the middle of the unique record, but, we will see it as an avoidable a part of grid cloud computing and performance computing. To anything degree, compared with general community carrier, cloud computing is easy to extend and has an easy management fashion. Cloud does not only certainly acquire the computer aid, however, but also

provides a control approach and may offer offerings for a huge wide variety of users on the identical time. This present day, there may be no filed of the facts middle wherein visualization is not applied [25]. It has ended up a critical device and makes the provider capability higher. Because the storage and computing potential of the server cluster is an excessive amount of, we don't want to buy servers, what we want is to feature a digital machine running at the server. If the cluster is massive sufficient, the request of including server may have a useless impact, after which we are able to shop the fee that needs to be spent on buying new servers. Simultaneously, cloud computing comes up with effective supports for SAAS (software program as a provider) [26]. It amalgamates all the organizations that yield comparable services at the net in order that users can evaluate and pick service companies. Cloud computing yields reliable and secure information storage middle, yields huge opportunity for internet utility, affords boundless area for storing and handling statistics, presents a robust computing capacity for customers to finalize all sorts of utility. Within the destiny, the pc may additionally only be used for connecting the internet to put in force offerings primarily based on cloud computing. Users' habit of the usage of a pc may exchange absolutely, from computer targeted offerings to internet centered offerings. Within the destiny, we only are seeking for a pocketbook personal pc or a mobile cellphone, then we are able to accomplish what we want thru net offerings including large duties inclusive of supercomputing. So quit-person is the actual owner of cloud computing. The want of software of cloud computing is to bring together all of the resources, and grant everyone can use it. A few of the most primary importance, cloud computing is to utilize software program and facts of the internet [27].

2.1.2. Characteristics of cloud computing

Ultra-large-scale

Cloud is simply too big in its length. The cloud of Google has owned over one million servers. Even cloud of Amazon, IBM, Yahoo, and Microsoft owns servers which are counted in loads of lots. Business enterprise has also many servers which might be counted in loads. Cloud provides the customers' computing potential.

Virtualization

Cloud computing we could the person get service being everywhere, thru any type of terminal. The assets it demanded come from the cloud in preference to a visible entity. You could

accomplish whatever you need via internet service using a personal computer or cellular smartphone. Users can get access to it safely through an easy manner, anytime, anywhere. This permits users to perform a challenge this is unable to be finalized in a single laptop. There special styles of virtualization based totally on the styles of hardware and application used.

A type of virtualization called hardware vitalization or hardware-assisted or server virtualization run at the concept that a separate impartial a part of a hardware or a physical server, is probably composed of many smaller hardware parts or servers, necessarily joining multiple physical servers into digital servers that run on a single number one physical server. Each small server can host a virtual machine, but, the complete cluster of servers is taken into consideration as a single device by something technique inquiring for the hardware. The hardware aid provision is performed through the hypervisor. The principal advantage is improved processing energy due to accelerated hardware usage and alertness uptime. There also are three subcategories of hardware virtualization inclusive of complete virtualization, partial virtualization, and Paravirtualization. Software program virtualization consists of the introduction of operation of many digital environments at the device that serves as a bunch. It produces a pc machine which is ready with hardware that shall we the guest OS to run. as an example, it enables you to run the Android running system on a bunch machine that normally the use of the Microsoft home windows operating the machine, the usage of the equal hardware because the host machine does. Software program virtualization itself divided into subtypes such as OS virtualization, application virtualization, and provider virtualization.

Physical reminiscence from one of a kind servers is blended together into one shareable virtualized memory. It affords the benefit of an enlarged contiguous running memory. You could not be new to this, as a few running gadgets including Microsoft windows working system we could a portion of your storage disk to paintings as an extension of your RAM. Memory visualization itself divided into distinctive subtypes which include application-degree manipulate and running machine degree manage.

Bodily storage devices that are many in quantity are grouped collectively, which after that appear as one garage tool. This substances special benefits which includes homogenization of garage across storage devices of any ability and speeds, decreased downtime, load balancing and higher

optimization of overall performance and pace. Dividing your difficult force into many partitions is an example of reminiscence virtualization. Block virtualization and record virtualize.

Information virtualization allows you to control records, as the information seems like a summary layer absolutely impartial of statistics structure and database structures. It reduces records to enter and formatting insects.

In community virtualization, many network branches can be created at the equal bodily network, which can also or won't is granted to talk with every different. This realizes restrict of report flow across networks and complements safety, and let’s better tracking and identification of statistics utilization which enables the network administrator to boom community insurance accurately. It also increases reliability as trouble in a single network does now not affect others, and the analysis is less complicated. There are two classes of network virtualization which might be internal network and outside network virtualization.

Computer virtualization lets the users’ running machine to be remotely saved on a server inside the statistics center, granting the consumer to then reaches their computer actually, from everywhere. Users that are searching for precise OS rather than home windows Server would require to have a virtual laptop. Advantages of computer virtualization are person mobility, effortless control of software program set up, updates, and patches.

High reliability

Cloud environments are regularly taken into consideration as offering a hundred percentage assured uptime, making them the primary among the maximum reliable services available within the web hosting industry. Its aid of a couple of servers, network and electricity redundancy, the aid (i.e. all server hardware must be maintained in a professional if it's fair to hold operational), and so on allows it to recognize the high reliability of the carrier. A consumer making use of cloud computing gets more dependable services than the only using the nearby laptop [29].

Versatility

Cloud computing does not simplest rely on a specific utility. It may also produce one of a kind sorts of applications which might be supported by using the cloud, and one cloud can aid diverse applications walking it simultaneously.

High extendibility

Cloud skills can be elastically supplied and launched, depending on the cases can be routine, to speedy increase its scale and decrease its size with demand. To the patron, the capacities exist for provisioning often look like as many as is possible and can be suited in any quantity at any time.in fashionable, the cloud can resize its scale dynamically and routinely relying on the incoming requests to satisfy customers' necessities.

On-demand service

Cloud is a massive shared aid that you should purchase primarily based for you want. Cloud is similar to water, electric, and gas that may be paid according to the quantity you utilized.

Extremely inexpensive

Due to the fact, the cloud's maximum capacity to suffer from the fault may be constructed through nodes that aren't too luxurious, the centralized control of cloud we could the organization want no longer undertake the control price of the information center that increases at an excessive rate. The flexibility can growth the usage price of the existing resources as compared with the antique gadget, so customers can absolutely use the low-value benefit. Nowadays, you may spend just a few hundred greenbacks and a few days to carry out an assignment that you ought to do it dropping thousands of greenbacks and some numerous months before.

2.1.3. Cloud computing service model

Cloud computing has three fashions for turning in services. These are Infrastructure as service (IAAS), software program as a provider (SAAS), and Platform as a provider (PAAS).

Infrastructure as a Service (IAAS)

Infrastructure as a provider is handiest one tenant cloud layer in which cloud computing dealer's owned sources are best shared among shrunk clients where the price is in step with the amount of utilization. This to an outstanding extent reduces the demand for an excessive amount of preliminary funding in computing hardware such as servers, networking devices, and processing strength. except, they aid varying levels of monetary and useful flexibility not exist in internal records facilities or with facet by side services, for the reason that computing sources can be brought or freed a great deal more speedy and fee-successfully than in an internal facts center or

with a facet by aspect provider [30]. Infrastructure as a service and different related offerings have enabled startups and different agencies consciousness on their valuable abilities without being concerned a good deal about the provisioning and control of the infrastructure. Infrastructure as a provider absolutely abstracted the hardware underneath it and granted customers to utilize infrastructure as a service without troubling something about the fundamental complexities. The cloud has a very exciting cost proposition in terms of value, but 'out of the box' Infrastructure as service simplest come up with basic safety consisting of perimeter firewall, load balancing, and many others. And packages going into the cloud will call for higher ranges of safety given at the host

Platform as a Service (PaaS)

Platform as a Service (PaaS) is a collection of software and development tools stored on the provider's servers. It is a single layer at the top of Infrastructure as a Service on the stack and condenses away all things up to the operating system, middleware, etc. This gives an integrated collection of developer surroundings that a developer will use to create their applications while not having any suggestion concerning what's happening below the service. It provides developers with a service that offers a full software system development lifecycle management, from attending to style to developing applications to readying to testing to maintenance. Every single issue else is abstracted aloof from the "view" of the developers. PaaS cloud layer functions like infrastructure as service, however, it provides an extra level of 'rented' practicality. End users exploitation the platform as a service services transfer even additional quantity of prices from capital investment to operational expenditure, however, ought to acknowledge the extra constraints and maybe some extent of lock-in caused by the additional practicality layers [31]. The employment of virtual machines acts as a catalyst within the platform as a service layer in cloud computing. Virtual machines should be unbroken safe against malicious attacks like cloud malware. For that reason, keeping the integrity of applications and powerfully implementing correct authentication checks at the time of the transfer of knowledge across the complete networking channels is crucial.

Software as a Service (SAAS)

SaaS could be a software system a software delivery model during which applications are held on by a vendor or service supplier and provided to customers through the network, particularly

the net. The software system as a service is changing into the associate progressively widespread delivery model as underlying technologies that support internet services and service-oriented design (SOA) totally developed and new biological process approaches become the desired one. The software system as a service is additionally oftentimes connected to a pay as you go subscription allowing model.

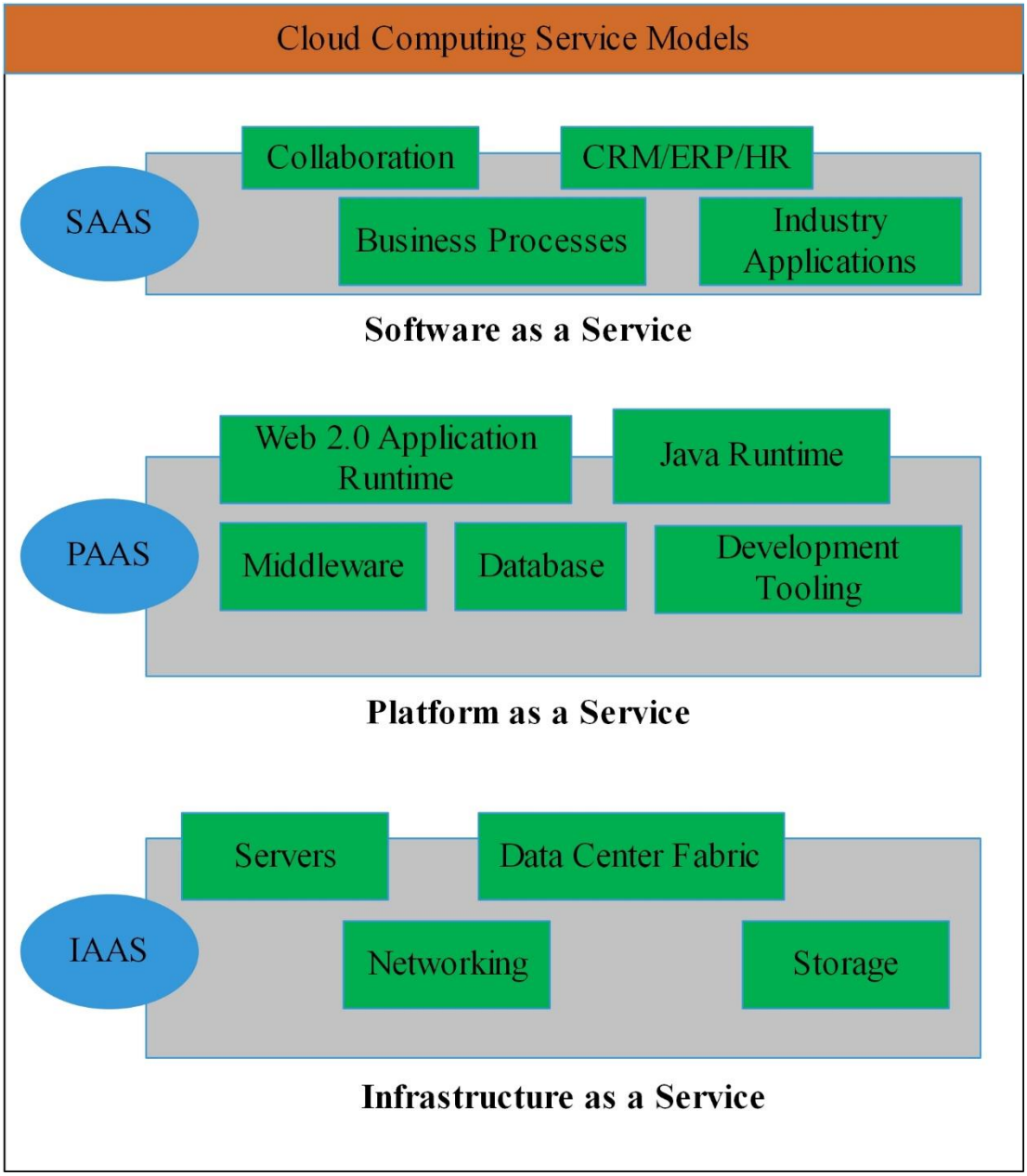


Figure 2. Cloud computing service delivery models [32]

For now, the broadband carrier has grown to be regularly on the market to consumer access from several geographical regions around the arena. software program gadget as a provider unremarkably enforced to deliver enterprise software program practicality to company customers at an espresso worthwhile belongings those customers advantage the identical edges of commercially authorized, internally operated software program system and not using a connected satisfactory of installation, management, steering, licensing, and excessive preliminary price. The design of software machine as a carrier based in general programs is unambiguously designed to aid several parallel customers at the equal time. SaaS programs are acquired thru packages over the net consequently net browser protection is surprisingly important. Data protection officials can need to offer some notion to absolutely special strategies of securing software program machine as a carrier application. internet services (WS) safety, lengthy language (XML) mystery writing, Secure Socket Layer (SSL), and present decisions which are employed in acting protection against the information transmitted via the net [33].

2.1.4. Cloud computing deployment models

There exist three cloud computing deployment models that are private, public, and hybrid cloud models.

Private cloud

The personal cloud is these days evolved term that a few vendors have lately used to inform of offerings that comply with cloud computing on personal networks. It's miles established within a agencies' inner organization statistics middle. With a private cloud, scalable assets and digital applications presented by means of the cloud dealer are blended together and unfastened for cloud customers to proportion and use. It varies from the public cloud in that entire the cloud resources and packages are managed by means of the corporations itself, the identical to net capability. Utilization on the private cloud may be noticeably secure than that of the public cloud as a result of its exact internal vulnerability. Best the enterprise and designed stakeholders may have get admission to work on a specific cloud [34].

Public cloud

Public cloud clarifies cloud computing in the traditional mainstream element, whereby sources are dynamically assigned on a first-class-grained, self-service basis via the internet, the use of net applications/net offerings, from an off-website online third-celebration supplier who stocks

resources and bills on an excellent-grained application computing basis. It mainly relies upon on pay in line with the utilization model, much like a pay as you go strength metering device that's dynamic enough to provide for spikes in want for cloud optimization [35]. Public clouds aren't as relaxed as other cloud fashions for the reason that it places an extra burden of ensuring all applications and information reached on the public cloud are not liable to malicious attacks.

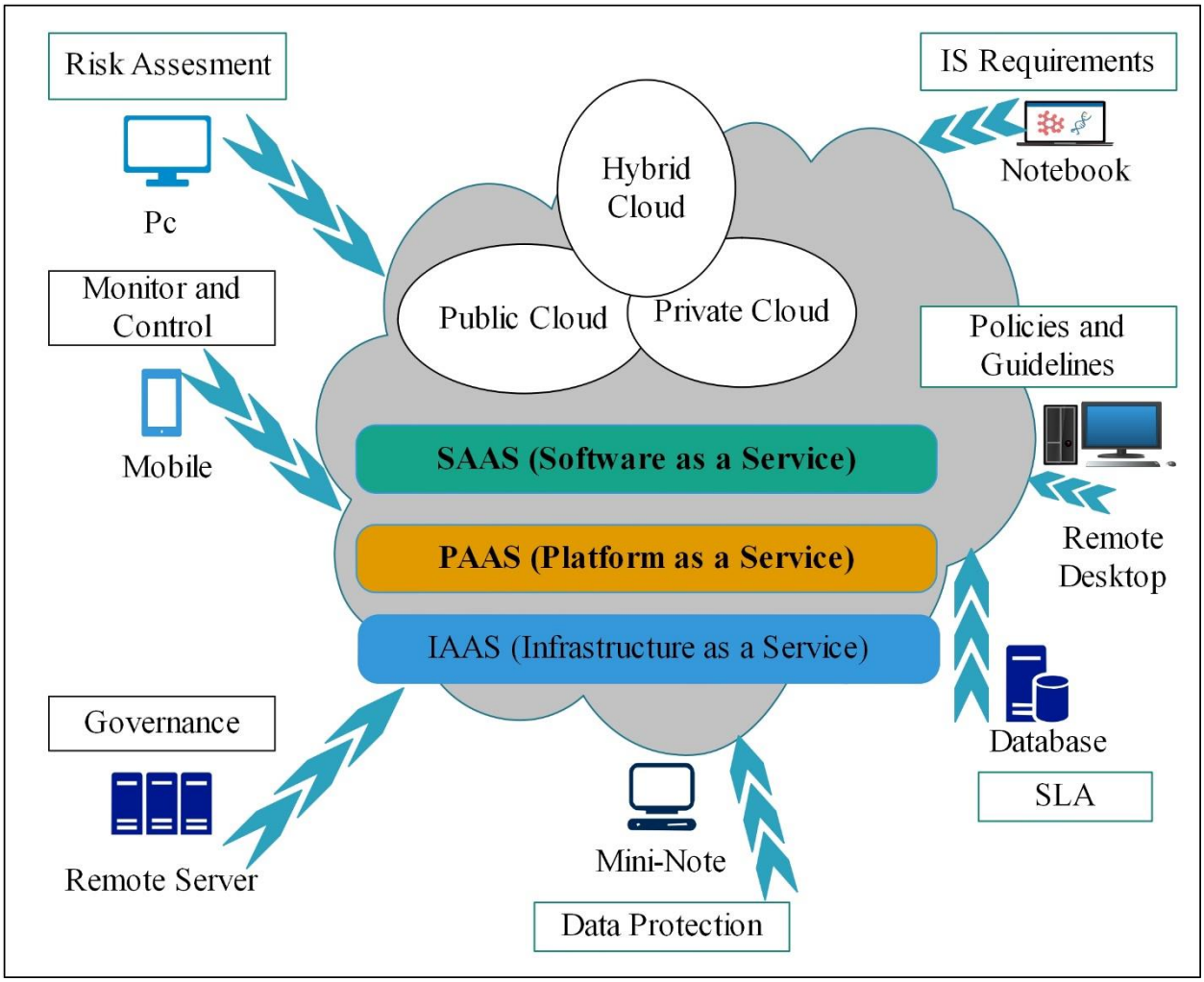


Figure 3. Cloud computing deployment models [36]

Hybrid cloud

Hybrid cloud is a private cloud connected to one or greater than outside cloud offerings, centrally managed, provisioned as one unit, and confined with the aid of a comfy community [37]. It gives digital statistics generation (IT) solutions through a combination of each public and personal clouds. A hybrid cloud offers greater comfy control of the information and applications and

permits diverse parties to attain statistics over the net. It additionally has an open architecture that we could interfaces which has other control structures. Hybrid cloud can describe configuration joining a neighborhood tool, along with a Plug laptop which has cloud services. It may additionally describe configurations becoming a member of virtual and physical, collocated belongings –for instance, a mainly virtualized surroundings that desire physical servers, routers, or different hardware inclusive of community equipment acting as a firewall or junk mail clear out.

2.1.5. Cloud computing storage and data utilization

Cloud storage is one kind of use of cloud computing. With the cloud storage, data is outsourced on many third-party servers, instead of on the dedicated servers utilized in traditional interconnected data storage. While storing data, the user sees a virtual server that is, it exists as if the data is loaded in a particular place with a specific name. However, that location does not exist in point of fact. It's simply a name location wont to reference virtual house sculpted out of the cloud. In reality, the owner's information may well be loaded on anybody or a lot of computers wont to build the cloud [38].

The actual storage place may even vary from day to day or maybe minute to minute because the cloud controls offered space for storing during a versatile manner. However, although the place is virtual, the user sees a static place for his/her information and might truly management his/her space for storing as if it were connected to his own notebook computer (PC).

At its most basic level, a cloud storage system desires simply one information server connected to the net. A user sends copies of his/her files via the net to the information server, which successively records the knowledge.

When a user must retrieve the file, he/she reaches the information server via a web-based interface. A method of reaching the information center is thru a keyword search. Several search systems were developed to form users ready to access their file of interest from the remote cloud being all over connected to the net. Mathematician search system, single keyword search system, and multi keyword search theme are samples of the search mechanisms developed to assist users' access and retrieve their files of interest from the cloud. The server then either returns the files

back to the user or lets the owner access and perform a dynamic operation on his/her files on the server itself.

Keyword search benefits for cloud computing

The benefits of cloud computing ar emotional information owner from the burden, upon demand service, no matter location and etc. the purchasers ought to retrieve solely information files that are of their interest. To grant this facility, one in every of the most effective manner is to specifically retrieve the files victimization keyword-based search instead of providing the full info no matter users' interest. Such search ways that rely upon the keyword provides solely the files that the purchasers would like google, yahoo, Wikipedia, and etc. the normal coding ways produce AN index for every keyword and this index is connected with the files. This methodology of building AN index for every keyword is termed secure trapdoors. However, it solely operates for actual keyword match that successively limits the users to conduct keyword search that isn't acceptable for cloud computing. It's identified that clients' input may not be the same as existed keywords like yahoo and yahoo. So, to higher search results accuracy still on improving user looking out expertise, it's to boot crucial for such ranking mechanism to support multiple keywords search with ranking practicality, jointly keyword search oft provides a way too coarse result. As a typical follow indicated by today's internet search systems like Google search, users may tend to input a group of keywords instead of one because the indicator of their search interest to retrieve the foremost connected information.

Adopting a keyword search for cloud computing

Using keyword search techniques for reaching information keep on the remote cloud started an extended time past for the aim of reducing the burden of accessing files that meet users' interest from the massive quantity of files that are counted in millions and billions keep within the remote cloud storage. Therefore implementing or applying search techniques to the current big quantity of information desires several advanced technologies like techniques for locating the file index that best matches with the user submitted search terms. Several researchers were operating so as to form the search over the cloud storage economical and improved than ever. Nowadays, several of the cloud-based sites like google, yahoo, and Amazon, etc. are using higher search mechanism for the aim of enabling users to access their files of interest expeditiously and effectively over the net.

Challenges of adopting keyword search in cloud computing

Although several recent works on keyword search over matter documents, for example, hypertext mark-up language documents, are designed, the present internet search engines (e.g., Google) often yields an inventory of individual pages as results of the search request and can't mix info from several interconnected pages answer keyword queries meaningfully. Within the event that there exist no pages that contain all the keywords, they're going to yield pages with a number of the input keywords stratified by similarity. Despite the chance that two or a lot of interconnected pages embrace all the keywords, already offered techniques unable of group action pages into one relevant and purposeful answer.

2.2. Related works

Privacy-retaining search over encrypted information and searchable coding techniques are deeply studied in recent years. There exist three basic models within the search over encrypted information schemes [39]. The primary one is that the merchant technique. During this model, the information loaded on the server isn't personal, however, the shopper desires to perform a look revealing no info on the information reached. Though this model isn't required to look over encrypted information, it's still a major privacy-retaining search model since the user will get a dataset item revealing no info concerning the accessed files or documents. Personal info retrieval (PIR) protocols yield an answer for this model [40]. The second model is that the load or store and moveable feast system, wherever a user will implement a look over the information that's encrypted beneath the owner’s public key. This model is a lot of pleasure for secure email applications wherever the receiver’s public secret is identified to the sender. Public key coding escorted by keyword looking out (PEKS) methodology for this model is ab initio designed by Boneh et al [41]. Many sequent enhancements on the general public key coding with keyword search methodology are designed in [39], each for single and conjunctive keyword search environments. The third model is that the storage system that's not public (i.e., info outsourcing scheme), wherever a data owner moves his/her data to a foreign server. During this model, the information is often within the encrypted type or in plain type. A number of the granted users will then perform a look operation on the encrypted information unseaworthy no sensitive info concerning the search request terms to the remote server.

In this work, the storage system model that's not personal and solely think about the occasion, wherever sensitive information is stirred to a foreign server in AN encrypted type is taken into account. Connected works for this theme are often analyzed in two main groups: single keyword search and multiple keyword searches. During a single keyword search user solely will submit a question that contains just one keyword to retrieve his/her file of interests. However, in multiple keyword searches user will send a look question that contains multiple keywords so as to urge the foremost relevant files of his/her interest.

2.2.1. Single keyword search

Research within the field of looking out over encrypted cloud information ab initio declared with single keyword search that restricts users to look for the file of his/her interest only mistreatment the only keyword in the search term.

D. X. Song, D. Wagner, and A. Perrig [9] were the primary to look at the ways for keyword search over encrypted information that are kept on the remote cloud. The author starts with the concept for storing a set of plaintext documents on information storage server like email servers and file servers in an encrypted format to deduce security and privacy issues. The work comes up with a scientific technique that permits indexed search on keep encrypted information unseaworthy no any sensitive data to the untrusted remote server. This technique is unable to support high service needs like looking out expertise and system usability. Additionally, to it, these work solely enable one keyword search to access and retrieve a file of interest, which successively is low as several users may submit multiple keywords instead of single keyword search as a hunt request input. Therefore, it's obligatory to develop search systems that support high service needs for keyword search over encrypted information.

C. Wang, N. Cao, K. Ren, and W. Lou [10] examined the matter of secure ordered keyword search over encrypted cloud information. It is an improvement of the work of Song et al. The author studied the applied mathematics count approach that embeds the similarity score of every file or document at the time of the institution of the searchable index before moving the encrypted assortment of the document to the remote cloud. The work includes the planning of one keyword searchable cryptography system with ranking criteria rely upon the keyword frequency that's of nice importance in returning best-matching documents. The planned search system conjointly solves process overhead, information and keyword privacy, and minimum

communication. During this technique, the information owner builds an index together with the keyword frequency depends on similarity scores for documents/files. The designed theme permits users to send a call for participation to the cloud server with the choice key as trapdoor along the personal key. The cloud server searches the index with scores and sends encrypted document or file consistent with hierarchical manner. However, this work is unable to perform multiple keywords because it is intended just for one keyword search theme. There's a bit overhead at the time index building.

Goh et al. [11] designed an index design that accomplished security. The secure index structure enabled an index associated with a keyword k to be checked in Q (1) time if k was a part of the searchable index. During this technique is was conjointly accomplished that no extra data may be extracted from the index until a sound trapdoor is yielded. The trapdoors are secure as a result of that they're designed with a secret key. Linguistics security against adaptively chosen keyword attack was their security model and that they conjointly designed IND-CKA secure index construction by the name (Z-IDX) that was economical as is employed Bloom filters at the same time with pseudo-random functions. The designed ways may be utilized for applications like constructing encrypted audit logs and different information question system wherever privacy ought to be accomplished. This work is additionally unable to support multiple keyword searches over the encrypted information because it is essentially designed for one keyword search system. Consequently, it's not economical and secured to a high degree.

C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou [12] designed a secure ordered keyword search over encrypted cloud information. hierarchical search system intensifies system usability to a good extent because it provides back the matching documents during a ranked manner rely upon some relevant criteria that are keyword frequency, as a result of that creating the search over cloud information a lot of easy.

D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano [13] were the primary to get searchable cryptography construction within the public configuration. In their strategy owner uses the general public key to load information on the cloud storage but for looking out the files of interest personal secret is obligatory. They came up with the concept of public key cryptography together with keyword search and established many constructions.

All the work that is analyzed on top solely has the power to conduct one keyword search. But, in some case of search over encrypted cloud information, the number of the loaded dataset is often normal big in size and single keyword search can mechanically come back an extravagant number of matches within which most of the resulted files are going to be tangential for the user. The most effective answer to the current drawback is multi keyword search. Multiple keyword searches let several constraints within the search request question and create the user capable of accessing solely the foremost relevant information.

2.2.2. Multi-keyword search

Multiple keyword searches could be a looking out a technique that permits users to incorporate over one keyword in their search question to induce the foremost applicable results.

N. Cao, C. Wang, M. Li, K. Ren, and W. Lou [14] advised a way that lets multiple keywords hierarchical search over encrypted cloud information. The authors introduced multiple keywords ordered search technique, wherever they utilized the principle of “coordinate matching” that want get the similarity between a multiple keyword search question and encrypted files loaded on to the cloud server. Notwithstanding, their index structure uses a binary illustration of keep file terms and for that reason, the hierarchical search doesn't differentiate files that contain the next range of continual terms than files that contain a lower number of repeated terms. During this technique, the information owner has got to send a regular key that is employed in index generation to all or any granted users. This causes high information measure which ends in high communication overhead. Besides, this work desires keyword fields within the index. This suggests that the user ought to grasp a listing all out there valid keywords and their position as obligatory data to provide a question. This thought might not be applicable in several cases. Their work is inefficient as a result of matrix operation operations of sq. matrices wherever the amount of rows is within the manner of some thousands.

C Orencik, E Savas [15] introduced a way that has represented and glued the matter of effective but safe and sound rank keyword search over encrypted cloud information. hierarchical search considerably strengthens system usability by giving back the matching documents during a ranked manner consistent with sure necessary criteria that is keyword frequency and consequently creating one step nearer towards smart consumption closing from secure information hosting services in cloud computing. In their work, the difficult drawback of privacy

enabled and economical multiple keyword searches over outsourced encrypted cloud information (MRSE) are represented and glued, and that they, in addition, established a group of severe privacy needs for such a protected cloud information usage system to become a reality. Their planned ordering technique proves to be economical to come back enormously relevant files related to user-submitted search terms. However, their work is applicable just for one information owner. Their planned technique is additionally not decent enough to permits the information owner to perform a dynamic operation like change and deletion files from the cloud.

Ning Cao et al [16] designed a hunt that's for a familiar cipher-text model and background model over encrypted information giving minimum computation and small communication overhead. The coordinate matching is chosen for multiple keyword search theme. They utilized the scalar product to quantitatively access the price of similarity for ranking documents. Their planned work that is multiple keyword searches over encrypted cloud information has tiny variance σ which incapacitate keyword privacy. Multiple keywords hierarchical search (MRSE) for familiar cipher theme might yield two non-identical trapdoors that imprecise the privacy leak drawback of trapdoor unlinkability which could incapacitate the keyword privacy. In addition, the uprightness of the ordering isn't confirmed in multiple keywords hierarchical search (MRSE). During this technique extra feature like an activity a dynamic operation on keeps encrypted cloud information isn't acceptable.

Zhihua Xia [17] designed a way that suggests a tree-based search model over encrypted cloud information, which has multiple keywords ordered search. The vector area theme and term frequency (TF) and inverse document frequency (IDF) scheme are along utilized within the building of index and creation of search question to yield multiple keywords ordered search result to achieve high search potency outputs, the author builds a tree-based structure and designed a Greedy Depth-first search rule rely upon this index tree to achieve higher potency than linear search model. As a consequence of this special structure of the tree-based index, the designed search theme can flexibly deliver the goods sub-linear search time and might effectively add an economical manner. The k-nearest neighbor (kNN) rule is employed to encipher the index and question vectors, and as a result, prove correct similarity score calculation between encrypted documents and question terms. The searchable cryptography model developed during this work permits the consumer to store the encrypted information to the cloud

and perform a keyword search over encrypted text domain. During this work, the cloud service suppliers (CSPs) that store the data for owners might access owners and users sensitive information with none authorization. This, in turn, breaks information privacy and security principle that ought to be accomplished in cloud data storage. The planned system doesn't support a feature like multiple information owner and there's high information measure that causes computation and communication overhead.

This study focus on improving the search accuracy and more concentrated on reducing a little bit time for looking the desired data in a big size of cloud data by applying a search theme that supports a user query which includes many key terms to the cloud server. The proposed work also utilizes the feature of advanced encryption standard algorithm such as minimum time for encrypting and decrypting files as well as for providing strong privacy through encryption using a 256 key size.

CHAPTER THREE

3. METHODOLOGY

3.1. Data collection method

3.1.1. Literature review

The first most step in this work begins by specifying the general area of the study followed by reading journals, papers, area-specific professional websites, and surveys regarding the identified area in order to make the area of study more specific. Following the specification of the study area, related works from secondary data sources are deeply studied. These data sources are digital libraries like google scholar, IEEE, research gate, and others. These sources are used for getting related works regarding the identified specific study area. Papers, journals, surveys, and publications from these sources are used for analysis and reviewed to study in details what the existing systems look like and how it works and also to get what was done and what was not. The study of these works is used for providing much information about the existing works and their limitations which is left for future works. In order to get all these information from the previously done works, great attention and time are given without getting bored of reading them to end until the core idea of the work is found.

3.1.2. Internet and websites

Bibliographies and internet search were used to progressively identify further relevant sources. This yielded a large number of resources which are filtered according to relevance and importance to the proposed system in this work. Official and non-official reports, as well as library approach, are also used to get any important information or data that is helpful for this proposed work.

3.2. Design method

3.2.1. Lucidchart

Lucidchart is an indispensable visual productivity ground which helps everyone understand and exchange ideas, informations, and processes with comprehensibility. With this subconscious, cloud-based solution, everyone who ever interested in using Lucidchart can learn to think visually and cooperate in real time at the time of building flowcharts, mockups, UML diagrams, and e.t. Lucidchart is used in more than 175 countries by more than 9 million users, inclusive of

Comcast, national acrostics and space administration, streaming services, and online providers. Starting from the establishment of Utah-depend firms in 2010, it is continually show growth by almost a 100 percent each year and has obtained numerous prizes for its profit earning jobs and a good discipline in their job atmosphere. Place to that, lucid-chart make available the changes to user accounts at the start of every month as to keep improving their program whenever you utilize it. Lucid-chart encompasses every libraries which are standardized that customer could expect from some of the programs that alternative choices of Visio, these encompasses flow-charts, swim-lanes, entity relationship diagram, wire-frames, Vienna charts, mind-drawing, and additionally designed libraries. Even though you are unable to found library that you demanded, you always start your own by taking in files with scale vector graphics format into an attitude behave library. Lucid-chart is most of the time examined to Visio which is of Microsoft. Two of them are private standard designing tools. As the best alternative of Visio SaaS providing, lucid-chart grants users to work on the same document at the same time with their friends with the help of any newly developed browser. Visualizing a complicated plan, series of activities, or system is much simple in lucid-chart than a Visio which is a property of Microsoft and E-draw max.

Because of all these its advantages over other diagramming tools, Lucidchart is appeared to be the best diagramming tool for this work and every design in this work is designed by Lucidchart.

3.3. Experimental method

3.3.1. Term Frequency-Inverted Document Frequency (TF-IDF)

The frequency of the term and frequency of the inverted document is a type of data mining technique which is used for the purpose of organizing files. It is very important to describe the dissimilarity between Frequency of the term and frequency of the inverted document and sentiment examination. Even though two of them are considered as a file grouping way, their aim is completely not the same. On one side, sentiment examination trying to group files into feelings such as positive and negative. On the other side, Frequency of the term and frequency of the inverted document groups files into groups inside the files themselves. This would come up with insight into what the examinations are about, rather than if the writer was feel good or feel bad. If we examined product examination data from an electronic commerce location which are exchanging computer spare parts for money, we would come to an end with groups of files about laptop, keyboard, desktop, etc. we would earn a big size of data about the kinds of deliveries that

had been given, however, never attempts to gain what user guesses about these items. Even though, the group of instruction is relevant in that they categorize text, the results of each yield us unique details

This algorithm is helpful whenever you have a file set, specifically a one with big size, which expected to be grouped. It is uniquely accurate because you should not need to train a scheme ahead of time and it will on itself take into account for dissimilarity in the sizes of files.

Think of a big size corporate websites which tolerate hundreds of thousands of customer-contributed blog posts. Depending on the tags connected to every single blog post, the item will present on listing pages on many different parts of the website. Though the writes were capable of tagging very things himself on manual-based at the time they wrote the details, in many scenarios they choose not to, and appropriately too many blog posts are not grouped. Empirics indicate that only little number of customers can have time to make something appear as tags on blog posts and aid with a grouping of what have been there on posts and re-check, making voluntary firms that cannot be continued. Such file is a good utilize scenario for Frequency of the term and frequency of the inverted document as a result of that it can make tags appear on the blog posts and aid them by making them appear at the correct part of the blog. Best of all, no need for a new trainee to tag them on their own going through them. Immediate loading of the algorithm could traverse through file set and order them appropriately making them convenient.

The frequency of the term and frequency of the inverted document calculates the amount that will represent the need of term inside a fil. It accomplishes this by contrasting and comparing the number of terms used inside that file as oppose to the whole combination of files. The need for that term continues growing as the term frequently used in a single file itself which intern called frequency of the term in a file. It is must to face difficulty when a file contains the same term more than one. That’s why Frequency of the term and frequency of the inverted document also hides this value by the quantity of term existence in the entire files which mean frequency of the inverted document.

Number of the existence of term t = (Number of times term t appears in a file) / (Complete number of terms in the file)

Number of the existence of inverted document of term t = log_e (Complete number of files / Number of files with term t in it).

Result = TF * IDF

Recently, a bag of visual words with TF-IDF (term-frequency – inverse-document frequency) weight [42] has verified to be a really no-hit approach for image and explicit object retrieval, even on giant corpora [43, 44].

Because Term Frequency Inverted Document Frequency (TF-IDF), categorizes documents into classes within the documents themselves, that is best suit for this work, instead of categorizing documents supported the writers opinion (such as positive or negative) opposition linguistics analysis, TF-IDF is chosen for coordinate matching, to show results of user question in a very stratified manner, during this work.

3.3.2. Advanced Encryption Standard (AES)

Advanced secret writing commonplace (AES) may be a new secret writing technique steered by National Institute of commonplaces and Technology (NIST) to switch encryption Standard (DES) in 2001. AES algorithmic rule will settle for any combination of information (128 bits) and a key length of 128, 192, and 256 bits. The algorithmic rule is introduced to as AES-128, AES-192, or AES-256, hopping on the key size. Throughout encryption-decryption method, AES technique goes through ten rounds for 128-bit keys, twelve rounds for 192-bet keys, and fourteen rounds for 256-bit keys for delivering final encrypted text or to achieve the first plain-text [45]. AES lets a 128-bit knowledge length that may be divided into four elementary operational blocks. These blocks are recognized as an associate array of bytes and regular as a matrix of the order of 4*4 that's known as the state. For each secret writing and cryptography, the cipher starts with associate AddRoundKey section. However, before reaching the ultimate spherical, this yield goes through 9 major rounds throughout every one of these rounds four transformations are accomplished. These transformations are sub-bytes, shift-rows, mix-columns, and add the spherical key. Within the final spherical, there's no mix-column transformation [46]. Figure four manifest the method of AES. AES cryptography is that the reverse method of secret writing. It uses inverse functions like inverse substitute bytes, inverse

shift rows, and inverse combine columns. Each spherical of AES is controlled by the subsequent transformations [47]:

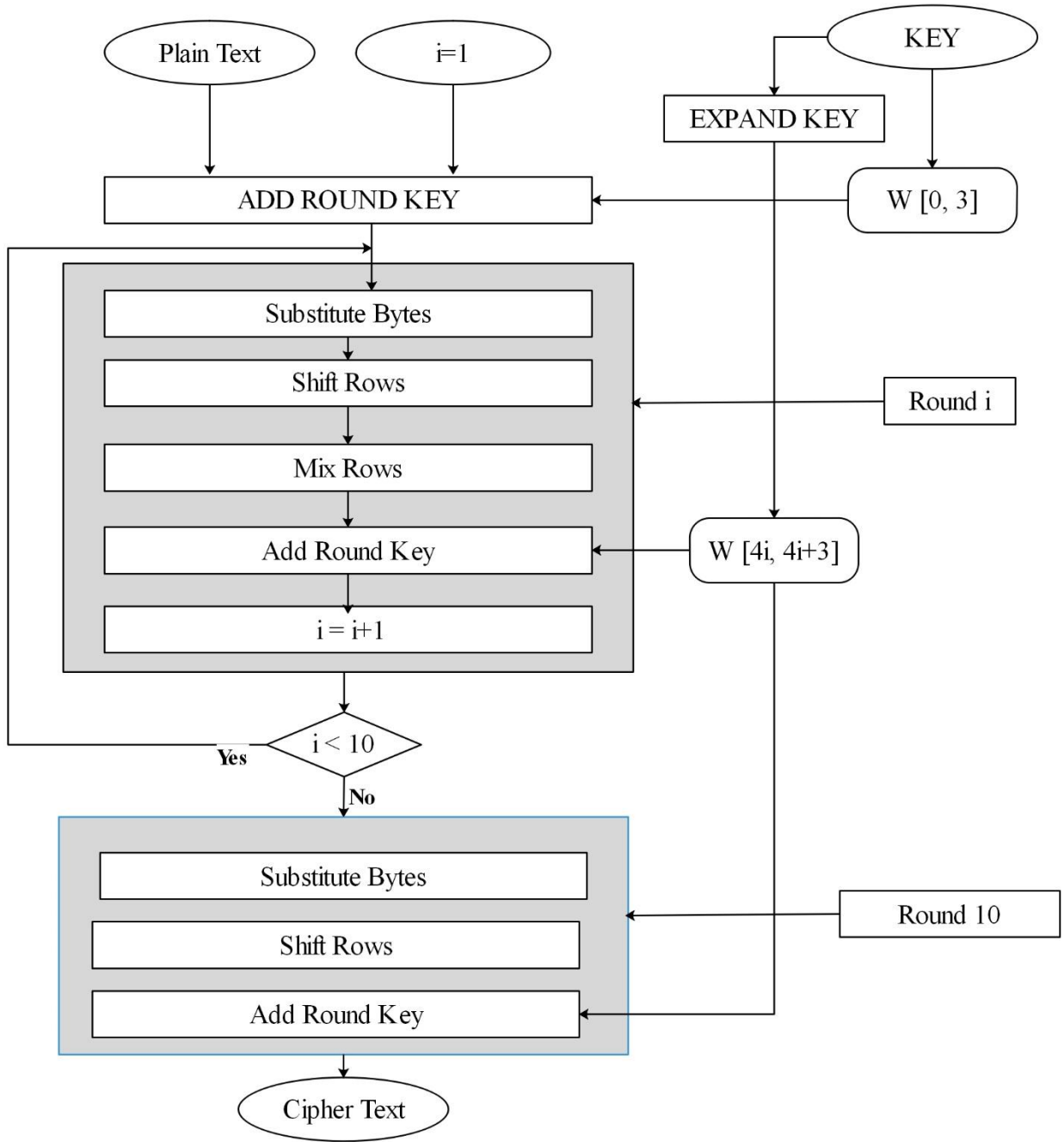


Figure 4. Advanced Encryption Standard (AES) Process [48]

Substitute computer memory unit transformation

AES accommodates 128-bit knowledge block, which suggests each one of the info blocks has sixteen bytes. In sub-byte transformation, every computer memory unit (8-bit) of an information

block is regenerate to associate other block victimization an 8-bit substitution box that is thought as Rijndael Substitution Box (S-box).

Shift Rows transformation

It is a straightforward byte transposition, the bytes in the last three rows of the state, depending upon the row whereabouts, are cyclically shifted. For 2nd row, the 1-byte circular left shift is accomplished. For the 3rd and 4th row, 2-byte and 3-byte left circular left shifts are accomplished respectively.

Mix columns transformation

This round is identical to a matrix multiplication of each Column of the states. A fix matrix is multiplied to every one of the column vectors. In this task, the bytes are treated as polynomials rather than numbers.

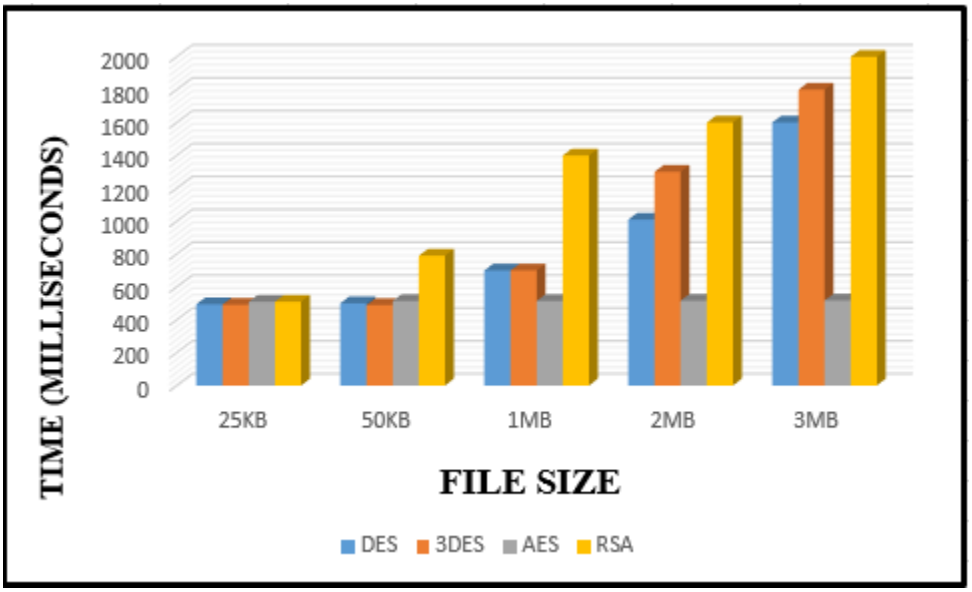


Figure 5. Encryption time vs. File size for DES, 3DES, AES, and RSA.

Add round key transformation

It is a bit-by-bit XOR between the 128 bits of current state and 128 bits of the round key. This transformation is the inverse of itself.

Evaluation parameters for selecting the right encryption technique

Encryption time

The time needed to change plaintext to cipher text is encryption time. Encryption time depends upon the key size and plaintext block size. Encryption time effects performance of the system. Encryption time must be less to make the system fast and responsive.

Decryption time

The time to get back plaintext from ciphertext is called decryption time. The decryption time should be less similar to encryption time to make the system responsive and fast. Decryption time effects performance of the system.

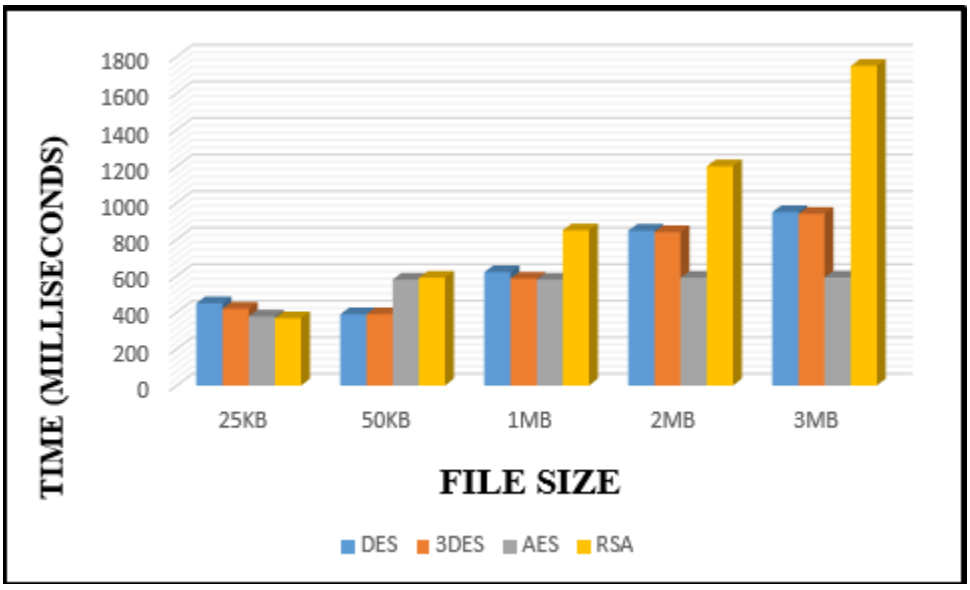


Figure 6. Decryption time vs. File size for DES, 3DES, AES, and RSA.

Memory used

Different encryption techniques need different memory size for implementation. This memory requirement depends on the number of tasks to be done by the algorithm, key size utilized, initialization vectors utilized and kind of tasks. The memory utilized effects the cost of the system. It is necessary that the memory needed should be as small as possible.

Table 1. Comparison of memory usage.

Algorithm	**Memory used (KB)**
DES	18.2
3DES	20.7
AES	14.7

RSA	31.5

Avalanche effect

In cryptography, an attribute known as diffusion reflects the cryptographic strength of an algorithm. If there is a small difference in input the output varies significantly. This is also known as called the avalanche effect. Avalanche effect is calculated using hamming distance. Hamming distance in the information hypothesis is a measure of dissimilarity. Hamming distance is found as the sum of bit by a bit XOR considering ASCI value, as it becomes simple to implement programmatically. A high degree of diffusion i.e. high avalanche effect is desired. Avalanche effect shows the performance of the cryptographic algorithm. Avalanche effect is calculated as the division of hamming distance by the file size.

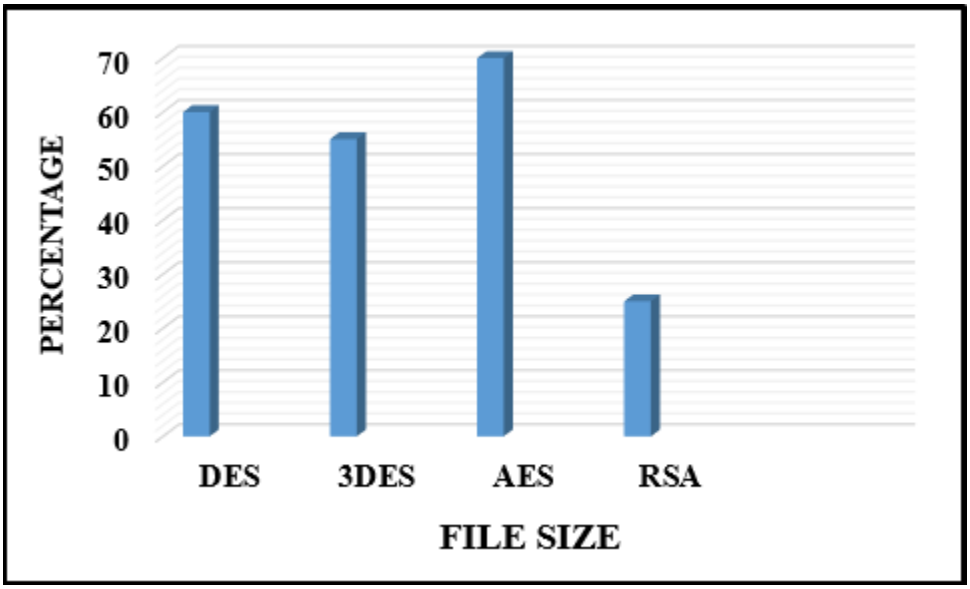


Figure 7. Avalanche effect for DES, 3DES, AES, and RSA.

Entropy

Randomness is a significant property in cryptographic processes because information should not be able to be guessed by an unauthorized user. Entropy is calculated of randomness in the information. It measures unpredictability in the information. In information security, security algorithms are required to yield high randomness in an encrypted message, so there is less or no dependency among key and ciphertext. With high randomness, the association between key and ciphertext becomes complex. This attribute is also called confusion. A high degree of confusion

is needed to make it difficult to guess to an unauthorized user. Entropy shows the performance of the cryptographic algorithm. Entropy is calculated using Shannon's formula.

Table 2. Average entropy values for DES, 3DES, AES, and RSA.

	DES	3DES	AES	RSA
Average entropy per byte encryption	2.9477	2.9477	3.84024	3.0958

Figure 5 shows that RSA takes the very best time for secret writing, however, AES uses the smallest amount time for secret writing, being quickest. Figure six shows that everyone algorithms need less time for secret writing than decipherment, RSA uses highest for decipherment, however, AES utilizes the smallest amount time for decipherment, being quickest. Figure seven shows that AES reveals the very best avalanche result, being sturdy insecurity whereas RSA reveals the smallest amount avalanche result. Table one shows that AES wants least memory house whereas RSA consumes highest memory house. Table a pair of manifests that AES scores highest average entropy per computer memory unit secret writing creating the output less liable to AN aggressor.

In general, as a result of these benefits of AES over the opposite techniques, AES is chosen for secret writing and decipherment during this work.

3.3.3. Merkle Tree Algorithm

Merkle tree is tree [49] in the course of which every non-leaf node is tagged with the hash of the labels of its youngster nodes. Hash bushes let cost-efficient and comfortable verification of the contents of massive knowledge structures. It's hash really worth is calculated from the concatenation of its left and right child hash really worth and every one leaf nodes save the hash of statistics blocks, in this documents/tokens hash well worth is hold. It makes use of the proper binary tree of left and right child. a perfect binary tree of depth h holds 2^h leaves. In alternative phrases, if range holds n tokens then the Merkle tree represents log (n) ranges. Hash map keep the (key, price) attempt for all tokens. Tree traversing is completed with a hash key of each node. Every node incorporates a hash key and key-word are searched with a hash key that is the picture of the linked set and aspect. It cuts off the tree in step with the question string. Tree hash carry out haphazardly selects the range stage and in line with that, it ranks the node inside the tree, all documents/tokens present in leaf nodes.

In this work, the Merkle tree algorithmic program is employed as a probe algorithmic program as an effect of it is comparatively cheap and powerful in retrieving files from an oversized amount of records. It's short in comparison to the extremely used DFS algorithmic program.

3.4. Simulation method

Simulation isn't something new in the realm of PCs. Recreation programming depends on the way toward displaying a genuine wonder with a lot of scientific recipes. It is, basically, a program that enables the client to watch an activity through reenactment without really playing out that task.

CloudSim

CloudSim is presumably the most well-known and prevalent test system for cloud situations and parameters. It was created in the CLOUDS Laboratory, at the Computer Science and Software Engineering Department of the University of Melbourne. The CloudSim library is utilized for [50]:

- Large scale distributed computing at server farms
- Virtualized server has with adaptable approaches
- Support for demonstrating and recreation of substantial scale distributed computing server farms
- Support for demonstrating and recreation of virtualized server has, with adjustable approaches for provisioning host assets to VMs
- Support for demonstrating and reenactment of vitality mindful computational assets
- Support for demonstrating and recreation of server farm organize topologies and message-passing applications
- Support for demonstrating and reproduction of combined mists
- Support for dynamic addition of reproduction components, just as ceasing and continuing recreation
- Support for client characterized arrangements to apportion hosts to VMs, and approaches for designating host assets to VMs

CloudSim programming structure and its design parts. The significant constraint of CloudSim is the absence of a graphical UI (GUI). In any case, in spite of this, CloudSim is as yet utilized in colleges and the business for the reproduction of cloud-based calculations.

ICanCloud

ICanCloud is a cloud test system which depends on SIMCAN. In straightforward words, iCanCloud can be utilized to reenact huge capacity systems [51]. So as to educate the clients about the expenses iCanCloud can foresee the exchange off among execution and cost. It centers on arrangements which charge clients in a compensation as-you-go way. The current programming frameworks must be displayed physically it has a full graphical UI from which analyses can be planned and run. It likewise permits parallel execution of one investigation more than a few machines.

CDOSim

CDOSim represents Cloud Deployment Option (CDO) Simulator which recreates the reaction times, expenses and SLA infringement of a cloud sending. It is a test system worried about choices which takes choice about the choices that cloud supplier makes, explicit runtime adjustment methodologies, segments organization of virtual machine and the arrangement of its occurrences. Segment arrangement to virtual machine cases incorporates the likelihood of making new parts out of prior segments. Virtual machine occasion's arrangement is normally the case kind of virtual machines. CDOSim can reproduce cloud arrangements of programming frameworks that were figured out to KDM models. CDOSim has capacity to speak to the client's as opposed to the supplier's point of view. CDOSim is a test system that permits the incorporation of fine-grained models. CDOSim is best case for looking at runtime reconfiguration plans or for deciding the exchange off among expenses and execution. CDOSim is intended to address the real weaknesses of other existing cloud test systems, for example,

- Consequently situated towards the cloud client point of view as opposed to uncovering fine-grained internals of a cloud stage.
- Mitigates the cloud client's absence of learning and control concerning a cloud stage structure.
- Simulation is free of solid programming dialects for the situation fitting KDM extractors exist for a specific language.

- Workload profiles from generation observing information can be utilized to replay real client conduct for reproducing CDOs.

GreenCloud

GreenCloud is a complex parcel level test system for vitality mindful distributed computing server farms with an emphasis on cloud correspondences. It offers a point by point fine-grained displaying of the vitality devoured by the server farm IT gear, for example, figuring servers, organize switches, and correspondence joins [52]. GreenCloud can be utilized to create novel arrangements in observing, asset assignment, remaining task at hand booking just as streamlining of correspondence conventions and system frameworks. It is discharged under the General Public License Agreement and is an expansion of the notable NS2 arrange test system [52]. The Key highlights of this test system are:

- Focus on cloud systems administration and vitality mindfulness
- Simulation of CPU, memory, stockpiling and systems administration assets
- Independent vitality models for each kind of asset
- Support of virtualization and VM relocation
- Network-mindful asset assignment
- Complete TCP/IP usage
- User well-disposed GUI
- Open Source

TeachCloud

TeachCloud is a cloud test system which is explicitly worked for instructive purposes. TeachCloud has a basic graphical interface through which understudies and researchers can change a cloud's setup and perform straightforward tests. TeachCloud utilizes CloudSim as the essential plan stage and presents numerous new improvements over it, for example,

- Developing a GUI toolbox.
- Adding the cloud outstanding task at hand generator to the CloudSim test system.
- Adding new modules identified with SLA and BPM.
- Adding new cloud organize models, for example, VL2, BCube, Portland and DCell.
- Introducing an observing outlet for the majority of the cloud framework parts.

- Adding an activity module that empowers understudies to reconfigure the cloud framework and concentrate the effect of such
- Changes on the complete framework execution.

Cloud Analyst

CloudAnalyst is a CloudSim based Visual Modeler for Analyzing Cloud Computing Environments and Applications. It was created to mimic substantial scale Cloud applications with the motivation behind contemplating the conduct of such applications under different sending arrangements. CloudAnalyst helps engineers with experiences in how to disseminate applications among Cloud foundations and esteem included administrations, for example, streamlining of uses execution and suppliers approaching with the utilization of Service Brokers [53]. The advancement of this test system has been finished utilizing Java, Java Swing, CloudSim and SimJava and the featuring highlights are:

- Ease of Use
- Ability to characterize a recreation with a high level of configurability and adaptability
- Graphical Output
- Repeatability
- Ease of Extension

GroudSim

GroudSim is an occasion based test system that needs one recreation string for logical applications on matrix and cloud situations dependent on an adaptable reproduction free discrete-occasion center. It is chiefly focused on the IaaS, however it is effectively extendable to help extra models, for example, PaaS, DaaS and TaaS. The client to recreate their examinations from a similar situation utilized for genuine applications by incorporating GroudSim into the ASKALON condition. GroudSim gives a far reaching set of highlights for complex recreation situations, for example, basic occupation executions on rented registering assets, estimation of expenses, and foundation load on assets. Reenactments can be parameterized and are effectively extendable by likelihood appropriation bundles for disappointments which regularly happen in complex conditions. Trial results exhibit the improved versatility of GroudSim contrasted with a related procedure based methodology.

System Cloud

System Cloud is an all-inclusive adaptation of CloudSim and is planned with the capacity to actualize organize layer in CloudSim. It takes a BRITE document as info and produces a topological system. The topology documents contain the quantity of hubs alongside the different substances engaged with reenactment. In reproduction, every element is to be mapped to a solitary BRITE hub with the goal that organize CloudSim can work effectively. System CloudSim is generally used to recreate the conduct of system traffic in CloudSim.

DCSim

The Data Center Simulator (DCSim) fundamentally takes a shot at the IaaS layer and offers administrations to various inhabitants and focuses on virtualized server farm so as to make a recreation to test and improve server farm the executives systems. For the provisioning of registering assets Data focuses are ending up progressively well known. In DCSim, a server farm comprises of a lot of interconnected Hosts (physical machines), represented by a lot of Management Policies. Each host has a lot of Resource Managers that handle nearby asset allotment, a CPU Scheduler which chooses how VMs will execute, and a Power Model which models how much power is being devoured by the host anytime. Each host runs a lot of VMs, which thusly each run a solitary Application. DCSim bolsters Virtual Machine the board tasks, for example, VM live movement and replication as a matter of course. The asset needs of each VM in DCSim are driven powerfully by an Application, which changes the dimension of assets required by the VM to reproduce a genuine remaining burden. The Application class is conceptual and can be stretched out to actualize distinctive sorts of utilizations, however the essential application demonstrate executed in DCSim mirrors a constant, value-based outstanding task at hand, for example, a web server. Approaching work can be load adjusted between various application examples running in independent VMs, and multi-layered applications can likewise be displayed [54].

The proposed work is implemented in CloudSim with cloud analyst using NetBeans in java language.

The general activities followed all through these paintings are proven in discern 8 under. It indicates every activity and each step accompanied to finish this painting starting from

identifying the studies vicinity to the improvement of the prototype and documentation of the proposed gadget.

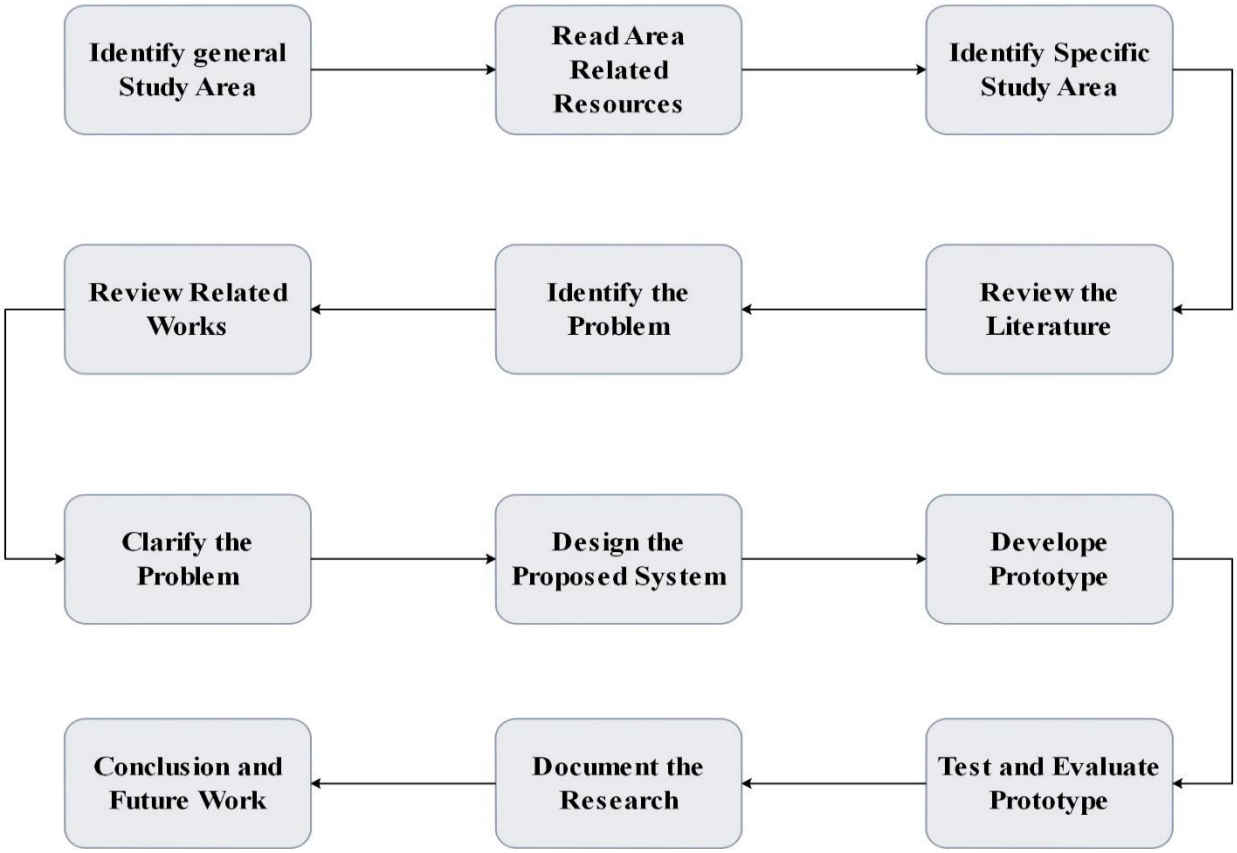


Figure 8. The architecture of the overall methodology followed

CHAPTER FOUR

4. THE PROPOSED METHOD

In this section, the proposed scheme, which especially includes the subsequent four phases: report Encryption, question technology, and report Retrieval are presented.

4.1. Overview

To stumble upon the necessities of green multi-keyword ranked seek in multiple records proprietor schemes, a novel PEMKSE mechanism is proposed in these paintings. Figure. 8 indicates the running techniques of PEMKSE.

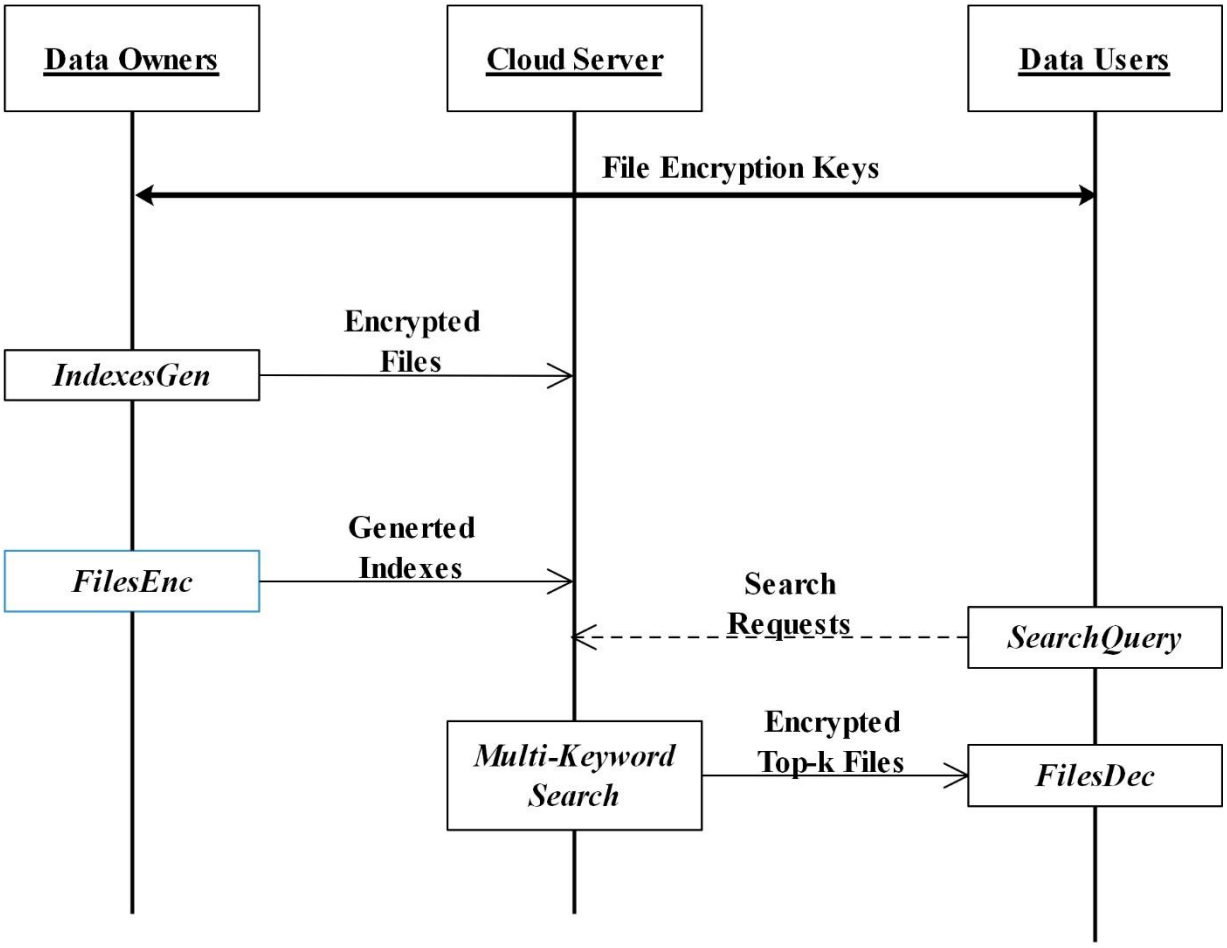


Figure 9. PEMKSE working processes

4.1.1. Data owners

FilesEnc makes use of the superior Encryption standard encryption algorithm to encrypt data owners' documents. IndexesGen builds the tree-primarily based index for every information proprietor’ files. Then information proprietors upload encrypted files and tree-based indexes to the cloud server.

In preferred, to preserve records non-public, the facts proprietor encrypts actual files using symmetric encryption. And he/she additionally produces the corresponding indexes. After that, he/she sends the encrypted documents and the corresponding indexes to the cloud server and sends the symmetric keys to records users.

4.1.2. Data Users

SearchQuery submits customers’ seek time period for his or her files of interest to the cloud server. FilesDec decrypts encrypted files with the proprietors’ mystery key.

In standard, the data consumer receives the symmetric that's the name of the game keys from the facts proprietor. After that he/she generates the trapdoor for his query which includes a set of textual key phrases and virtual key phrases. And he/she additionally produces a selection set. After that, he/she submits the trapdoor at the side of the range set to the cloud server. Sooner or later, after receiving the back files, he/she decrypts the documents with the received symmetric keys.

4.1.3. Cloud Server

Multi-key-word seek accepts the user seeks request and returns the top-okay outcomes in a ranked manner to the users.

The cloud server is an in-between entity that stores the encrypted documents and corresponding indexes that is sent from statistics proprietors. It offers search provider after receiving a trapdoor and a selection set from the guest user. It calculates the inner product of the index vector and the trapdoor for every each record to go back pinnacle-k documents with the very best scores on the idea that they meet all of the various calls for.

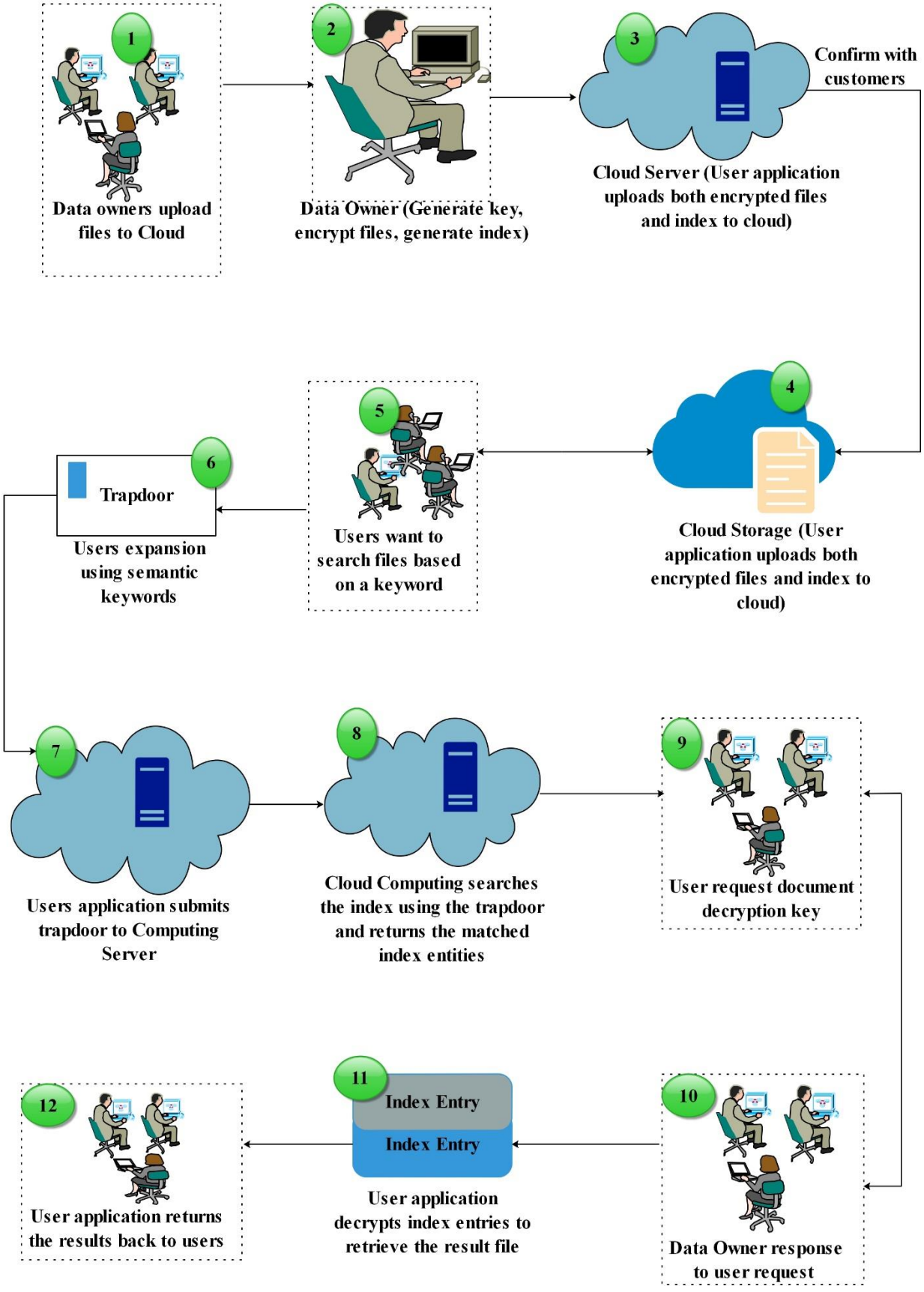

1
Data owners upload files to Cloud
2
Data Owner (Generate key, encrypt files, generate index)
3
Cloud Server (User application uploads both encrypted files and index to cloud)
Confirm with customers
4
Cloud Storage (User application uploads both encrypted files and index to cloud)
5
Users want to search files based on a keyword
6
Trapdoor
Users expansion using semantic keywords
7
Users application submits trapdoor to Computing Server
8
Cloud Computing searches the index using the trapdoor and returns the matched index entities
9
User request document decryption key
10
Data Owner response to user request
11
Index Entry
Index Entry
User application decrypts index entries to retrieve the result file
12
User application returns the results back to users

Figure 10. The architecture of the proposed system.

As displayed in figure 10 the proposed gadget composed of four entities: - facts owner, information user, and Cloud Server. Statistics pre-processing, encryption, decryption all venture are achieved via consumer application gadget module be composed of 3 modules as data preprocessing i.e. tokenization of phrases, construction of index tree required for looking, and question processing module. Records preprocessing module does diverse preprocessing functions on the complete statistics set and generate a list of tokenized phrases. Construction of an index tree module constructs index tree from which we need to search particular key-word through the use of term frequency values (TF) values of a keyword from every document. Query processing module manipulates the question terms and searches over the index tree. The similarity between the phrases and document is calculated and set ranked applicable files are returned to the user.

User registration: the consumer has to first sign in himself to the gadget to apply offerings from the cloud provider issuer. Registered user can then add documents to the gadget and also search desired documents from encrypted document collection deposited on the cloud carrier issuer. A list of registered users is preserved on the Cloud provider issuer to authenticate customers and yield carrier to them.

4.2. Document Encryption and Uploading

In the proposed device, a well-known AES set of rules is applied to encrypt the information record to be uploaded with the aid of the owner. Superior Encryption well known is a symmetric block cipher. It uses the alike key for each encryption and decryption. AES general states that the algorithm can most effective receive a block length of 128 bits and a choice of 3 keys -128, 192, 256 bits. On this work, 256 bits with 14 rounds were used. The system of uploading a record through the owner entails of generating the important thing, encryption of a document accompanied via uploading.

The algorithm for encryption and uploading is provided below:

Step 1: Data Owner login to the system

Step 2: Data owner uploads files.

Step 3: If the files' name is new

Generate file key for encryption

Encrypt a file using AES algorithm

Upload the file to the cloud

Else if filename already exists

Overwrite/Stop

Step 4: end

AES encryption algorithm 1

```
Algorithm Start (FileName, SecretKey)
{
        Browse files
        Enter SecretKey
}
Algorithm Encrypt ()
{
        Convert to State Array
        AddRoundKey()
        SubBytes()
        ShiftRows()
        MixColumns()
        Key Expansion
}
```

4.3. Multi-Keyword Search

In the proposed system, multi keyword search algorithm for searching multiple keywords in a single query is utilized. When a data user is interested in reading/viewing any file of his/her interest, he/she submits related multiple keywords. The keywords are divide into an array of strings. Every word in the String Array is compared with the first record and if it exists it is added to a search output list and the lists are displayed to the user.

The algorithm is given below:

Step 1: Input Multi-Keywords to the System

Step 2: Split the words into a String Array

Step 3: For every word in the String Array

    Compare the word with the first record (file)

        If Found

            Add this file to the Search Output List

        Else Continue

Step 4: End For

Step 5: Display the output list to the user

Algorithm 2 ranked search

```
for all documents Ri do
    Compare(level1 index of Ri , query index)
    j = 1
    while match do
        increment j
        Compare (levelj indices of Ri, query index)
    end while
    rank of Ri = highest level that matches with query index
end for
```

4.4. Decrypting and downloading file

After the output list is displayed to the requestor/user, the user has to seek the file from the data owner. If the data owner is interested to share the file, he/she sends a key to the requestor/user. After receiving this key, the requestor can download and decrypt the file using the key.

The algorithm is provided below:

Step 1: Data user requests a key to download the file

Step 2: If the key is valid

Download and decrypt the file

Else if the key is invalid

Error while downloading

Step 3: End If

Algorithm 3 file decryption

```
Algorithm Start (ReturnedResult, SecretKey)
{
    Enter SecretKey
}
Algorithm Decrypt()
{

    Convert to State Array

    AddRoundKey()
    InvSubBytes()
    InvShiftRows()
    InvMixColumns()
    Key Expansion

}
```

CHAPTER FIVE

5. RESULTS

5.1. Implementation results

In this section, the proposed method is extensively analyzed in order to reveal the efficiency and effectiveness of the scheme. The whole system is implemented by java language using a 64-bit Windows 7 operating system with Intel (R) Core (TM) I7 CPU @ 2.00GHz 2.60 GHz HP laptop as front end.

The evaluation criteria include precision and search time for searching using a Merkle tree search algorithm. In the area of information retrieval, precision is defined as the fraction of retrieved documents that are relevant to the user query. For instance, on a text search on a collection of documents, precision is the number of exact results divided by the number of all returned results.

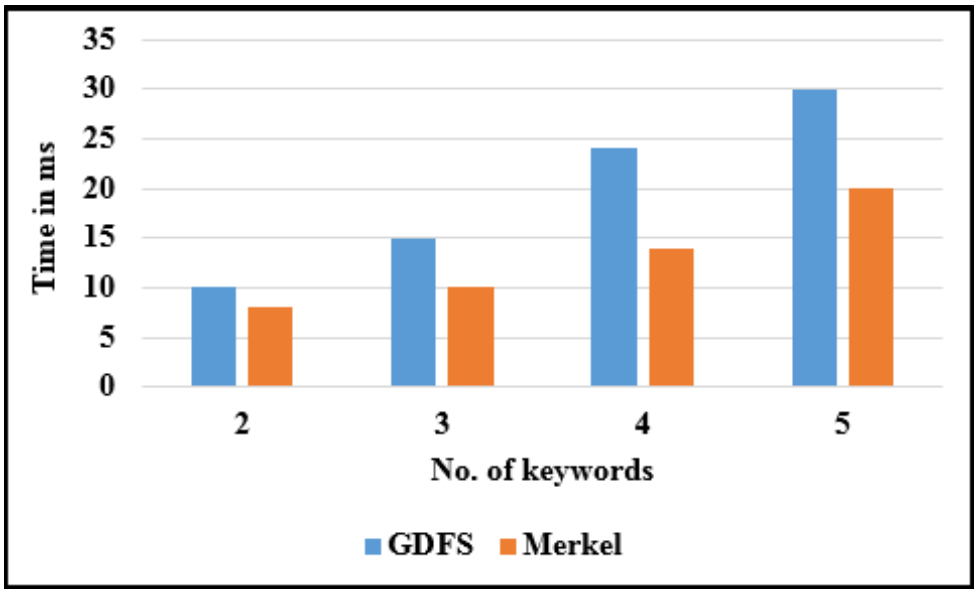


Figure 11. Searching time for keyword search in GDFS vs Merkel

During the keyword search process, Merkle Hash Tree calculate the hash code for the keyword and searches in the tree which is built from hashing, so the comparison will be easy and it gives O(log n) time complexity. In another hand, Greedy Depth First Search algorithm gives O (ωmlog n). Figure 11 reveals the particular number of query and the time it takes for searching. It shows

that the search technique used in previous works which are DFS and GDFS is not effective and efficient as much as Merkel tree which is used in this proposed method.

The screenshots of the results of the proposed system are shown below:

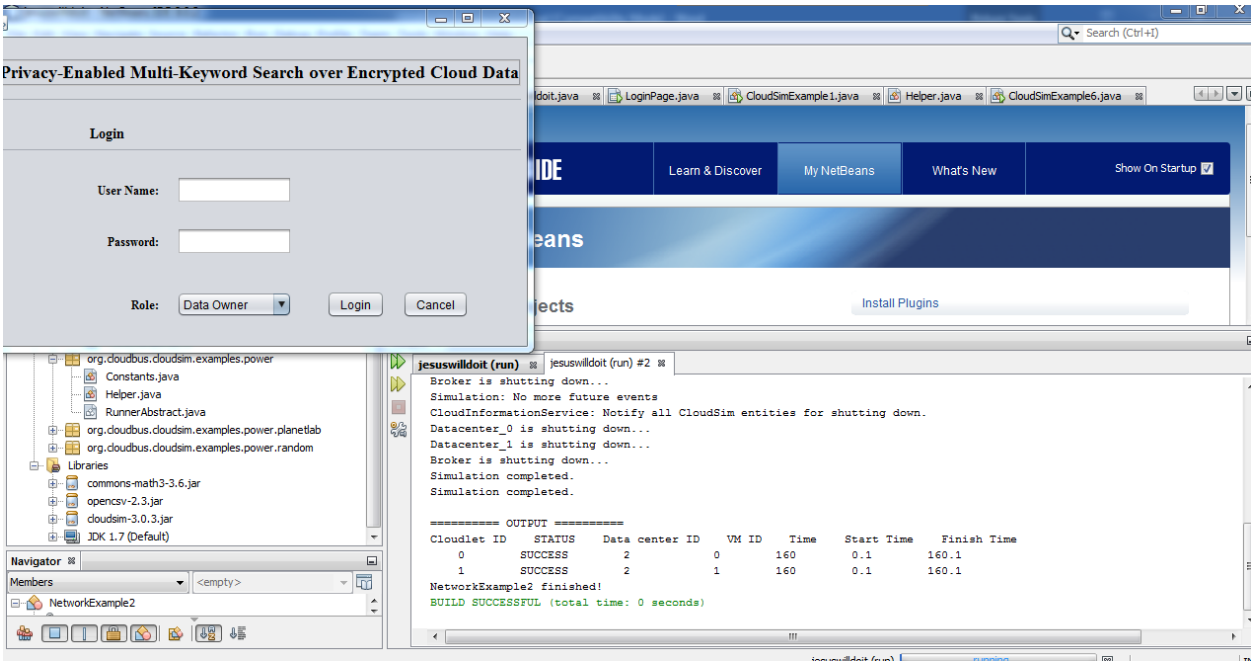
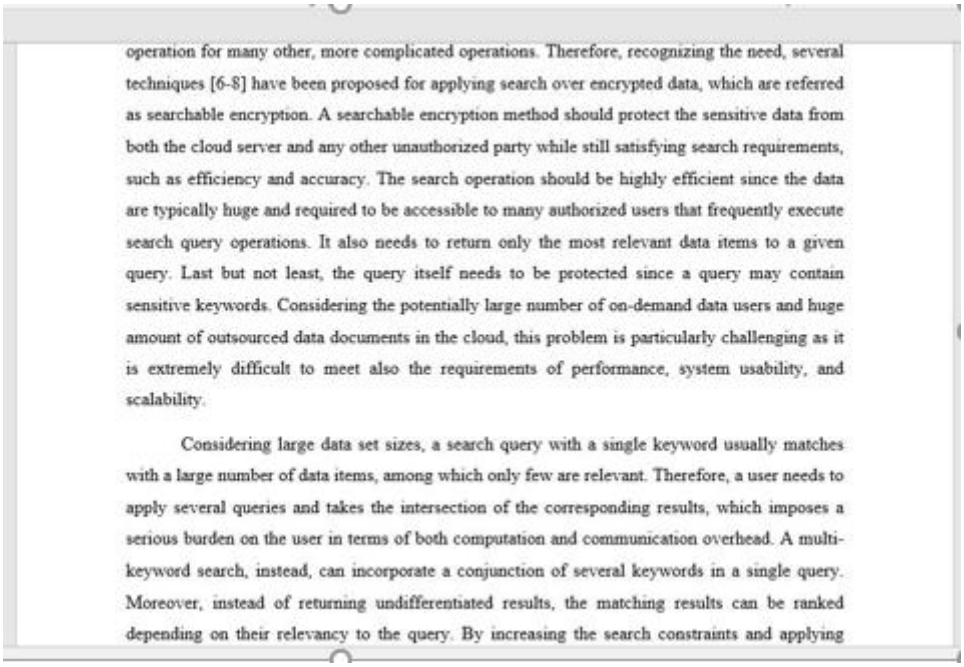


Figure 12. Data owner file upload page



operation for many other, more complicated operations. Therefore, recognizing the need, several techniques [6-8] have been proposed for applying search over encrypted data, which are referred as searchable encryption. A searchable encryption method should protect the sensitive data from both the cloud server and any other unauthorized party while still satisfying search requirements, such as efficiency and accuracy. The search operation should be highly efficient since the data are typically huge and required to be accessible to many authorized users that frequently execute search query operations. It also needs to return only the most relevant data items to a given query. Last but not least, the query itself needs to be protected since a query may contain sensitive keywords. Considering the potentially large number of on-demand data users and huge amount of outsourced data documents in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability, and scalability.

Considering large data set sizes, a search query with a single keyword usually matches with a large number of data items, among which only few are relevant. Therefore, a user needs to apply several queries and takes the intersection of the corresponding results, which imposes a serious burden on the user in terms of both computation and communication overhead. A multi-keyword search, instead, can incorporate a conjunction of several keywords in a single query. Moreover, instead of returning undifferentiated results, the matching results can be ranked depending on their relevancy to the query. By increasing the search constraints and applying

Figure 13. Document before encryption

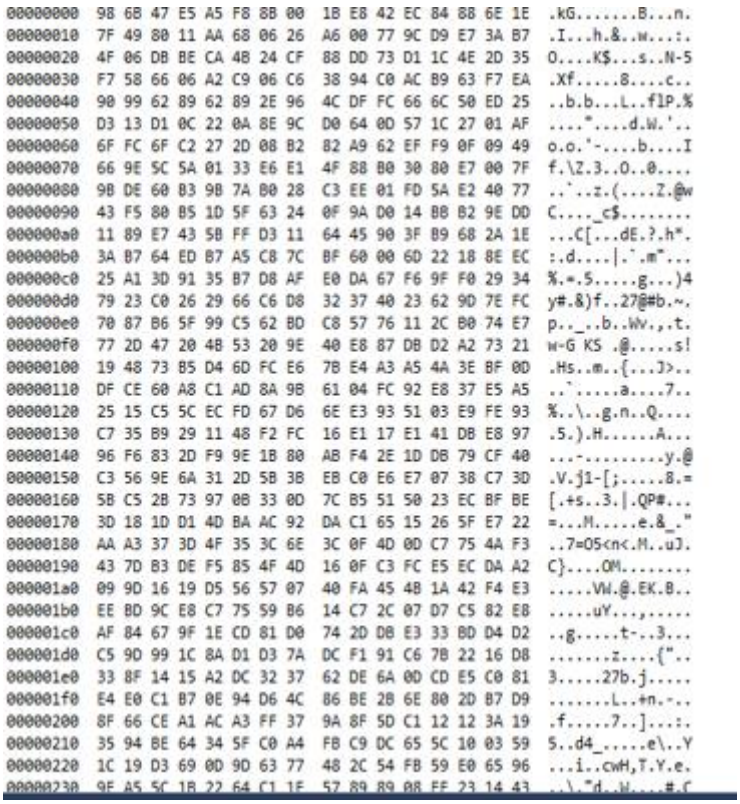


Figure 14. Document after encryption

As the proposed system uses AES as an encryption technique, the document privacy is strong compared to previous works which used other encryption algorithms for securing their data.

5.2. System testing

5.2.1. Unit testing

The inner software good judgment function correctly. This authorized with the aid of proving the proper inputs to the advanced prototype and getting the predicted output. All selection branches are tested and authorized with the correct drift of message. This test proves that after the right enter is given to the gadget the anticipated result is produced.

5.2.2. Integrated testing

The proposed machine is tested whether the complete device additives work together as one or not and it approves the integrity of the proposed gadget.

5.2.3. Functional testing

Purposeful checks offer systematic demonstrations that functions tested are available as precise by using the business and technical requirements, system documentation, and consumer manuals. Purposeful trying out is focused on the subsequent objects: valid input: recognized classes of valid input must be widely wide-spread. Invalid input: diagnosed training of invalid input has to be rejected. Capabilities: recognized functions need to be exercised. Output: recognized lessons of application outputs should be exercised. Structures/strategies: interfacing structures or methods must be invoked.

Organization and training of purposeful exams are focused on necessities, key features, or unique test instances. Similarly, systematic coverage pertaining to perceive commercial enterprise method flows; records fields, predefined tactics, and successive tactics should be taken into consideration for testing. Before purposeful trying out is entire, additional tests are diagnosed and the effective fee of present-day exams is determined.

Test objectives

- All discipline entries have to paintings well.
- Pages should be activated from the recognized link.
- The entry display, messages, and responses should not be delayed.

Features to be tested

- verify that the entries are of the best format
- No replica entries have to be allowed
- All links ought to take the consumer to the ideal page.

In trendy the proposed machine is approved to be correct and efficient as nicely as it meets the goal it becomes planned to gain.

CHAPTER SIX

6. FUTURE WORK AND CONCLUSION

6.1. Conclusion

The proposed work is focused on how the consumer can efficiently store their private files on a cloud while retaining the privacy of their documents and each time and anyplace vital they can seek and retrieve them by using sending a query inclusive of multiple key phrases to the cloud through internet-based total services. Privacy could be executed by way of encrypting the documents previous to outsourcing. In reaction to the user's question, the device will compare the keywords from question to the files the usage of "keyword-matching precept". Top-ranked files get retrieved so that you can include key phrases distinct by way of the person in a question. Checking the rank of the retrieved record can be completed with the aid of calculating how many documents involves the specified keywords on how regularly.

The work in this paper, define and remedy the problem of multi-keyword ranked seek over encrypted cloud statistics. Among diverse multi-key-word semantics, the efficient similarity measure of "coordinate matching," i.e., as many fits as viable, to efficiently document the relevance of outsourced files to the query keywords, and use "internal product similarity" to quantitatively verify such similarity degree is selected. For satisfying the venture of helping multi-keyword semantic without privacy breaches, a primary idea of PEMRSE using cozy inner product computation is proposed. After that, stepped forward PEMRSE schemes to acquire various stringent privateers requirements in specific risk fashions are given. a few further improvements of our ranked search mechanism, together with assisting extra search semantics, i.e., TF IDF, and dynamic facts operations additionally investigated. Thorough analysis investigating privateers and efficiency ensures of proposed schemes is granted, and experiments at the simulation show that the proposed schemes gift low overhead on both computation and verbal exchange.

6.2. Future work

Succeeding the modern-day studies, there are viable furtherance and present process efforts with a purpose to seem within the future work. first off, the personal aspect of the proposed system could be carried out on cellular gadgets jogging Android and iOS running systems for the reason

that capability software scheme envisions that users get right of entry to the statistics anywhere and every time. And secondly, the proposed technique might be examined on an actual dataset on the way to evaluate the overall performance of our rating technique with the rating techniques used in plain datasets that do not include any security or privacy-keeping strategies. Thirdly, the safety of the question, so as to be a tough difficulty desires to be advanced.

References

1 P. Mell and T. Grance, "*Draft NIST working definition of cloud computing,*" Online at http://csrc.nist.gov/groups/SNS/cloud-computing/index.html, 2010.

2 Cloud Security Alliance, "*Security guidance for critical areas of focus in cloud computing,*" Online at http://www.cloudsecurityalliance.org, 2009.

3 M. Bellare, A. Boldyreva, and A. ONeill, "*Deterministic and time-saving searchable encryption,*" in *Proc. of Crypto'07*, 2007, pp. 535–552.

4 Y. Chang and M. Mitzenmacher, "*Privacy-preserving keyword searches on remotely encrypted data,*" in *Proc. of ACNS'05*, 2005, pp. 442–455.

5 D. Boneh and B. Waters, "*Conjunctive, subset, and range queries on encrypted data*," in *Theory of Cryptography*, ser. LNCS, 2007, vol. 4392, pp. 535–554.

6 C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "*Secure ordered keyword search over encrypted cloud data,*" in *Proc. of ICDCS'10,* 2010, pp. 253–262.

7 A. Singhal, "*Modern information mining: A short overview*," IEEE Data Engineering Announcement, vol. 24, no. 4, pp. 35–43, 2001.

8 I. H. Witten, A. Moffat, and T. C. Bell, "*Managing gigabytes: Packing and indexing files and images*," Morgan Kaufmann Publishing, San Francisco, May 1999.

9 D. X. Song, D. Wagner, and A. Perrig, "*Practical methods for searches on encrypted data*". In happenings of the 2000 IEEE Symposium on Security and Privacy, Berkeley, CA, USA, May 2000.

10 C. Wang, N. Cao, K. Ren, and W. Lou' "*Allowing secure and efficient ordered keyword search overloaded cloud data*". IEEE Proceedings on Parallel and Distributed Systems, 23(8):1467–1479, 2012.

11 E.-J. Goh, "*Secure Indexes*", IACR Cryptology ePrint Archive, vol. 2003, pp. 216, 2003.

12 C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, "*Secure ordered Keyword Search over Encrypted Cloud Data*," in 30th International Convention on Distributed Computing

Systems, pp. 253-262, 2010.

13 D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, ”*Public key encryption with keyword search,*” in EUROCRYPT, pp. 506-522, 2004.

14 N. Cao, C. Wang, M. Li, K. Ren, and W. Lou’ “*Privacy-conserving multi-keyword ranked search over encrypted cloud data*”. In happenings of the 30th IEEE International Conference on Computer Communications, Shanghai, China, Apr. 2011.

15 C Orencik, E Savas, “*Efficient and Secure Ranked Multi-Keyword Search on Encrypted Cloud Data”*, IEEE PAPER, 2010.

16 Ning Cao et al.,” *Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data*", IEEE Transactions on parallel and distributed systems, vol. 25, no. 1, Jan 2014

17 Zhihua Xia, “*A Secure and Dynamic Multiple keywords ordered Search Techniques over Encrypted Cloud Data*", IEEE proceedings on parallel and distributed systems, vol: pp no: 99 the year 2015

18 N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, "Privacy-conserving multiple keywords ordered search over encrypted cloud data," in Proc. IEEE INFOCOM, Shanghai, China, 2011, pp. 829–837.

19 S. C. Yu, C. Wang, K. Ren, and W. J. Lou, “Realizing secure, scalable, and fine-grained data access control in cloud computing,” in Proc. IEEE INFOCOM, San Diego, CA, 2010, pp. 1–9.

20 http://searchcloudcomputing.techtarget.com/sDefinition/0,,sid201_gc i1287881,00.html

21 https://www.zdnet.com/article/what-is-cloud-computing-everything-you-need-to-know from public-and-private-cloud-to-software-as-a-service/

22 http://www.boingboing.net/2009/09/02/cloud-computing-skep.html

23 (U.S.) Nicholas. Carr, fresh Yan Yu, "IT is no longer Needed: the Internet great revolution of the high environment - cloud computing," The Big Switch: Re-wining the world from Edison to Google, CITIC Publishing House, October 2008 1-1

24 Ya-Qin Zhang, "*the future of computing in the cloud - Client*", The Economic Eyewitness reported, http://www.sina.com.cn, 2008 Nian 07 Yue 12 Ri 14:30

25 Wang Haopeng (Air Force Aviation University of Computer Teaching, Jilin, Changchun 130022, China); Liu strong (Air Force Air University, Research Department, Jilin, Changchun 130022, China), "*virtualization technology in the application of cloud computing"*, TP313.A ,1009-3044 (2008) 25-1554-01,2008 Year 25

26 http://www.emc.com/digital_universe, EMC 2008 Annual Overview Releasing the power of information

27 Ouyang Jing, "*how can the Cloud Computing be helpful*" - IBM Tivoli general manager Alfred Zollar interview, November 2008 11

28 https://ws680.nist.gov/publication/get_pdf.cfm?pub_id=909505, NIST Cloud Computing Reference Architecture, September 2011

29 http://hi.baidu.com/mc625263041/blog/item/22aeb1d0a492bd309a50276b.html, Cloud computing development status, 2009- 03-01 16:17

30 J. Brodkin. (2008, Jun.). "Gartner: Seven cloud-computing security risks." *Infoworld*, Available:http://www.infoworld.com/d/security-central/gartner-seven cloud-computing security- risks-853? Page=0, 1 [Mar. 13, 2009].

31 Global Netoptex Incorporated. "*Demystifying the cloud. Important opportunities, crucial choices*." pp4-14. Available: http://www.gni.com [Dec. 13, 2009].

32 M. Klems, A. Lenk, J. Nimis, T. Sandholm and S. Tai. "*What's Inside the Cloud? An Architectural Map of the Cloud topography*." *IEEE Xplore*, pp 23-31, Jun. 2009.

33 S. Subashini, and V. Kavitha. (2010) "*A survey on security problems in service delivery models of cloud computing*." Network Computing Application Jul. 2010.

34 S. Arnold (2009, Jul.). "*Cloud computing and the issue of privacy*." *KM World*, pp14-22. Available: www.kmworld.com [Aug. 19, 2009].

35 A Framework Computing Whitepaper. "*Public Cloud Computing: Translating IT*." Platform Computing, pp6, 2010

36 A Platform Computing Whitepaper. "*Enterprise Cloud Computing: Translating IT*." Platform Computing, pp6, 2010.

37 Global Netoptex Incorporated. "Demystifying the cloud. Important opportunities, crucial choices." pp4-14. Available: http://www.gni.com [Dec. 13, 2009].

38 Wassim Itani Ayman Kayssi Ali Chehab, "Privacy as a Service: Privacy-Aware Data Storage and Processing in Cloud Computing Architectures", Eighth IEEE International Proceedings on Dependable, 2009

39 Zhao, Y., Chen, X., Ma, H., Tang, Q., Zhu, H, "*A new trapdoor almost identical public key encryption with a keyword search."* J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl. **3**(1/2), 72–81 (2012)

40 Cachin C., Micali S., Stadler M, "*Computationally private information retrieval with polylogarithmic communication."* In: Lecture notes in computer science, pp. 402–414, Springer (1999)

41 Boneh, D., Crescenzo, G.D., Ostrovsky R., Persiano G.: Public key encryption with keyword explore. In: EUROCRYPT, pp. 506– 522 (2004)

42 J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In Proc. ICCV, 2003.

43 D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In Proc. CVPR, 2006.

44 J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In Proc. CVPR, 2007.

45 Mr. Gurjeevan Singh, Mr. Ashwani Singla and Mr. K S Sandha, "*Cryptography Algorithm Juxtaposition for Security Improvement in Wireless Intrusion Detection System*", International Journal of Multidisciplinary Research, Vol.1 Issue 4, pp. 143-151, August 2011.

46 William Stallings, "*Cryptography and Network Security: Fundamental and Practice*", Pearson Education/Prentice Hall, 5th Edition.

47 Akash Kumar Mandal, Chandra Parakash and Mrs. Archana Tiwari, "*Performance Judgement of Cryptographic Algorithms: DES and AES*", IEEE Students' Proceedings on Electrical, Electronics and Computer Science, pp. 1-5, 2012.

48 Feifei. L, Bin. Y, Ming Wang. T, and Marios. H, "Spatial Approximate String Search," IEEE Proceedings on Knowledge and Data Engineering, VOL. 25, NO. 6, JUNE 2013.

49 K. Ren, C. Wang, Q. Wang et al., "*Security issues for the public cloud*," IEEE Internet Computing, Vol. 16, No. 1, pp. 69–73, 2012

50 http://cloudsim-setup.blogspot.in/

51 Mohana S.J., Saroja M., Venkatachalam M., "Analysis and Comparison of Simulators to Evaluate the Performance of Cloud Environments", Journal of NanoScience and NanoTechnology, Vol 2, Issue 6, Spring Edition, pp 739-742, ISSN 2279 – 0381

52 http://greencloud.gforge.uni.lu/

53 Wickremasinghe B., Calheiros R.N., Buyya R., "CloudAnalyst: A CloudSim-based Visual Modeller for Analysing CloudComputing Environments and Applications",

54 Ostermann S., Plankensteiner K., Prodan R., and Fahringer T., "GroudSim: An Event Based Simulation Framework for Computational Grids and Clouds", M.R. Guarracino et al. (Eds.): Euro-Par 2010 Workshops, pp. 305–313, 2011. Springer- Verlag Berlin Heidelberg, 2011