**EFFICIENT REGULAR LANGUAGE SEARCH FOR SECURE CLOUD STORAGE**

A Project Report submitted in partial fulfillment of the degree of the   
Bachelor of Technology in Computer Science and Engineering

By

R. Priyanka 20C45A0513

D. Sneha 19C41A05H4

V. KavyaSri 19C41A05E5

Under the Guidance of

Mr.B.Samba raju

Assistant Professor



**Department of Computer Science and Engineering**

**Jayamukhi institute of technological Sciences**

**Narsampet, warangal-506 332**

**(Affiliated to JNTUH, Accredited by NAAC ‘A’ Grade)**

**2023**

**Jayamukhi institute of technological Sciences**

**Narsampet, warangal – 506 332**

**(Affiliated to JNTUH, Accredited by NAAC ‘A’ Grade)**

**CERTIFICATE**



This is to certify that the Project Report entitled **“EFFICIENT REGULAR LANGUAGE SEARCH FOR SECURE CLOUD STORAGE”** is a bona fide work of the students R.Priyanka, D.Sneha, V.KavyaSri bearing Roll No.s 20C45A0513, 19C41A05H4, 19C41A05E5 submitted in partial fulfillment of the requirements for the award of the degree of ***Bachelor of Technology*** in **Computer Science & Engineering** during the academic year **2022-23.**

**Guide Head of the Department**

**ACKNOWLEDGEMENT**

We express our immense gratitude and sincere thanks to **Prof. V. JANAKI**, principal, Jayamukhi Institute of Technological Sciences, Narsampet and Warangal for her constant and gracious guidance throughout the work.

We express our gratitude and sincere thanks to **Dr. P. SRINIVAS RAO**, Head of the Department Of CSE, Jayamukhi Institute of Technological Sciences for his constant motivation and support.

We are grateful to our Project Guide **Mr. B.Samba raju**, Assistant professor for his Valuable guidance and giving us the opportunity to do our work in a better manner.

We are also thankful to our Management for providing all the facilities in completing the project.

ABSTRACT

Cloud computing provides flexible data management and ubiquitous data access. However, the storage service provided by cloud server is not fully trusted by customers. Searchable encryption could simultaneously provide the functions of confidentiality protection and privacy-preserving data retrieval, which is a vital tool for secure storage. In this paper, we propose an efficient large universe regular language searchable encryption scheme for the cloud, which is privacy-preserving and secure against the off-line keyword guessing attack (KGA). A notable highlight of the proposal over other existing schemes is that it supports the regular language

encryption and deterministic finite automata (DFA) based data retrieval. The large universe construction ensures the extendability of the system, in which the symbol set does not need to be predefined. Multiple users are supported in the system, and the user could generate a DFA token using his own private key without interacting with the key generation center. Furthermore, the concrete scheme is efficient and formally proved secure in standard model. Extensive comparison and simulation show that this scheme has function and performance superior than other schemes.

**CONTENT:**

1. **INTRODUCTION 1-2**
   1. System requirements
2. **LITERATURE REVIEW 3-7**
3. **SYSTEM DESIGN 8-12**

3.1 System Architecture

3.2 Architecture diagram

3.3 Flow chart

3.4 Data flow diagram

3.5 Data flow levels

**4) MODULES DESCRIPTION 13**

**5)SYSTEM STUDY 14-17**

5.1 preliminary investigation

5.2 Feasibility analysis

5.3 Input design

5.4 Output design

**6)UML DIAGRAMS 18-19**

6.1 Use Case Diagram

6.2 Class Diagram

**7) SOFTWARE ENVIRONMENT 20-49**

**8) SYSTEM TESTING 50-53**

8.1 Unit testing

8.2 Integration testing

8.3 User Acceptance testing

**9)CONCLUSION 54**

**10)BIBILOGRAPHY 55-56**

1. **INTRODUCTION**

Cloud storage is an emerging model of storage to provide scalable, elastic and pay-as-you-use service to cloud computing users. For individual usage, the subscribers enjoy the freedom to access to their data anywhere, anytime with any device. When cloud storage is utilized by a group of users, it allows team members to synchronize and manage all shared documents. Moreover, it also saves the user a lot of capital investment of expensive storage equipments.

Cloud delivers convenience to the customers and at the same time arouses many security and privacy problems. Since the data are physically stored on the multiple servers of the cloud service provider, the customers cannot fully in charge of their data. They worry about the privacy of the stored documents since the server may be intruded by hacker or the data could be misused by the internal staff for commercial purpose. The customers prefer to adopt the encryption technology to protect the data confidentiality, which meanwhile arouses another problem: how to execute data retrieval on the large volume of ciphertext. It is almost unimaginable to ask the cloud subscriber to download all of their stored information and then decrypt and search on the recovered plaintext documents. No customer could tolerate the huge transmission overhead and the waiting time for the data retrieval result. Searchable encryption technology not only exerts encryption protection of the data, but also supports efficient search function without undermining the data privacy. The data user generates a token of the content that he wants to search using his private key. Receiving the token, the cloud server searches on the encrypted data without decrypting the ciphertext. The most important point is that the server learns nothing about the plaintext of the encrypted data nor the searched content during the data retrieval procedure. However, most of the available searchable encryption schemes only support some basic search patterns, such as single keyword search, conjunctive keyword search

and boolean search. Since the cloud computing is a fierce competition industry, it is of vital importance to provide good user experience. It is urgent to design novel searchable encryption schemes with expressive search pattern for cloud storage

**1.1 SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**Software Requirements:**

* Operating System - Windows XP
* Coding Language - Java/J2EE(JSP,Servlet)
* Front End - J2EE
* Back End - M

1. **LITERATURE REVIEW**

**Literature 1**

**Toward network-level efficiency for cloud storage services**

**Author : Li Z, Dai Y, Chen G, et al**

**LITERATURE DISCRIPTION :**

Cloud storage services such as Dropbox, Google Drive, and Microsoft OneDrive provide users with a convenient and reliable way to store and share data from anywhere, on any device, and at any time. The cornerstone of these services is the data synchronization (sync) operation which automatically maps the changes in users’ local filesystems to the cloud via a series of network communications in a timely manner. If not designed properly, however, the tremendous amount of data sync traffic can potentially cause (financial) pains to both service providers and users. This chapter addresses a simple yet critical question: Is the current data sync traffic of cloud storage services efficiently used? We first define a novel metric named TUE to quantify the Traffic U sage Efficiency of data synchronization. Based on both real-world traces and comprehensive experiments, we study and characterize the TUE of six widely used cloud storage services. Our results demonstrate that a considerable portion of the data sync traffic is in a sense wasteful, and can be effectively avoided or significantly reduced via carefully designed data sync mechanisms. Overall, our study of TUE of cloud storage services not only provides guidance for service providers to develop more efficient, traffic-economic services, but also helps users pick appropriate services that best fit their needs and budgets.

**Literature 2**

**Dynamic remote data auditing for securing big data storage in cloud computing**

**Author : Sookhak M, Gani A, KhanMK, et al**

**LITERATURE DISCRIPTION :**

Cloud computing has emerged as a new computing paradigm that offers great potential for storing data remotely. Presently, many organizations have reduced the burden of local data storage and maintenance by outsourcing data storage to the cloud. However, integrity and security of the outsourced data continues to be a matter of major concern for data owners due to the lack of control and physical possession over the data. To deal with this problem, researchers have proposed remote data auditing (RDA) techniques. However, the majority of existing RDA techniques is only applicable for static archived data and is not applicable for auditing or dynamically updating the outsourced data. They are also not applicable to big data storage because of the high computational overhead on the auditor. In this paper, we propose an efficient RDA technique based on algebraic signature properties for a cloud storage system that incurs minimum computational and communication costs. We also present the design of a new data structure-Divide and Conquer Table (DCT)-that can efficiently support dynamic data operations such as append, insert, modify, and delete. Our proposed data structure can be applied for large-scale data storage and will incur minimum computational cost. A comparison between our proposed method and other state-of-the-art RDA techniques shows that our method is secure and highly efficient in reducing the computational and communication costs on the server and the auditor.

**Literature 3**

**Privacy-preserving doubleprojection deep computation model with crowdsourcing on cloud for big data learning**

**Author : Zhang Q, Yang L T, Chen Z, Li P**

**LITERATURE DISCRIPTION :**

Recent years have witness a considerable advance of Internet of Things with the tremendous progress of communication theories and sensing technologies. A large number of data, usually referring to big data, have been generated from Internet of Things. In this paper, we present a double-projection deep computation model (DPDCM) for big data feature learning, which projects the raw input into two separate subspaces in the hidden layers to learn interacted features of big data by replacing the hidden layers of the conventional deep computation model (DCM) with double-projection layers. Furthermore, we devise a learning algorithm to train the double-projection deep computation model. Cloud computing is used to improve the training efficiency of the learning algorithm by crowdsourcing the data on cloud. To protect the private data, a privacy-preserving double-projection deep computation model (PPDPDCM) is proposed based on the BGV encryption scheme. Finally, experiments are carried on Animal-20 and NUS-WIDE-14 to estimate the performance of DPDCM and PPDPDCM by comparing with DCM. Results demonstrate that DPDCM achieves a higher classification accuracy than DCM. More importantly, PPDPDCM can effectively improve the efficiency for training parameters, proving its potential for big data feature learning.

**Literature 4**

**Two-Factor Data Security Protection Mechanism for Cloud Storage System**

**Author : Liu J K, Liang K, Susilo W, Liu J**

**LITERATURE DISCRIPTION :**

In this paper, we propose a two-factor data security protection mechanism with factor revocability for cloud storage system. Our system allows a sender to send an encrypted message to a receiver through a cloud storage server. The sender only needs to know the identity of the receiver but no other information (such as its public key or its certificate). The receiver needs to possess two things in order to decrypt the ciphertext. The first thing is his/her secret key stored in the computer. The second thing is a unique personal security device which connects to the computer. It is impossible to decrypt the ciphertext without either piece. More importantly, once the security device is stolen or lost, this device is revoked. It cannot be used to decrypt any ciphertext. This can be done by the cloud server which will immediately execute some algorithms to change the existing ciphertext to be un-decryptable by this device. This process is completely transparent to the sender. Furthermore, the cloud server cannot decrypt any ciphertext at any time. The security and efficiency analysis show that our system is not only secure but also practical.

**Literature 5**

**Privacy-Preserving and Regular Language Search Over Encrypted Cloud Data**

**Author : Liang K, Huang X, Guo F, Liu J K**

**LITERATURE DISCRIPTION :**

Using cloud-based storage service, users can remotely store their data to clouds but also enjoy the high quality data retrieval services, without the tedious and cumbersome local data storage and maintenance. However, the sole storage service cannot satisfy all desirable requirements of users. Over the last decade, privacy-preserving search over encrypted cloud data has been a meaningful and practical research topic for outsourced data security. The fact of remote cloud storage service that users cannot have full physical possession of their data makes the privacy data search a formidable mission. A naive solution is to delegate a trusted party to access the stored data and fulfill a search task. This, nevertheless, does not scale well in practice as the fully data access may easily yield harm for user privacy. To securely introduce an effective solution, we should guarantee the privacy of search contents, i.e., what a user wants to search, and return results, i.e., what a server returns to the user. Furthermore, we also need to guarantee privacy for the outsourced data, and bring no additional local search burden to user. In this paper, we design a novel privacy-preserving functional encryption-based search mechanism over encrypted cloud data. A major advantage of our new primitive compared with the existing public key based search systems is that it supports an extreme expressive search mode, regular language search. Our security and performance analysis show that the proposed system is provably secure and more efficient than some searchable systems with high expressiveness.

**Literature 6**

**Towards Achieving Data Security with the Cloud Computing Adoption Framework**

**Author : Chang V, Ramachandran M**

**LITERATURE DISCRIPTION :**

Offering real-time data security for petabytes of data is important for Cloud Computing. A recent survey on cloud security states that the security of users’ data has the highest priority as well as concern. We believe this can only be able to achieve with an approach that is systematic, adoptable and well-structured. Therefore, this paper has developed a framework known as Cloud Computing Adoption Framework (CCAF) which has been customized for securing cloud data. This paper explains the overview, rationale and components in the CCAF to protect data security. CCAF is illustrated by the system design based on the requirements and the implementation demonstrated by the CCAF multi-layered security. Since our Data Center has 10 petabytes of data, there is a huge task to provide real-time protection and quarantine. We use Business Process Modeling Notation (BPMN) to simulate how data is in use. The use of BPMN simulation allows us to evaluate the chosen security performances before actual implementation. Results show that the time to take control of security breach can take between 50 and 125 hours. This means that additional security is required to ensure all data is well-protected in the crucial 125 hours. This paper has also demonstrated that CCAF multi-layered security can protect data in real-time and it has three layers of security: 1) firewall and access control; 2) identity management and intrusion prevention and 3) convergent encryption. To validate CCAF, this paper has undertaken two sets of ethical-hacking experiments involved with penetration testing with 10,000 trojans and viruses. The CCAF multi-layered security can block 9,919 viruses and trojans which can be destroyed in seconds and the remaining ones can be quarantined or isolated. The experiments show although the percentage of blocking can decrease for continuous injection of viruses and trojans, 97.43% of them can be quarantined. Our CCAF multi-layered security has an average of 20% better performance than the single-layered approach which could only block 7,438 viruses and trojans. CCAF can be more effective when combined with BPMN simulation to evaluate security process and penetrating testing results.

**Literature 7**

**Adaptive E-learning using Deterministic Finite Automata**

**Author : Sarkat P, Kar C**

**LITERATURE DISCRIPTION :**

Adaptive E-learning refers to adapt the way of presentation of educational material according to the student’s needs. Understanding ability differs by student to student so the learning path should vary according to their understanding ability. Some students may understand by once some may needs more with different way. This paper represents a same topic with various approaches to the different classes of students with different understanding ability. The proposed approach in this paper is based on two concepts Deterministic Finite Automata (DFA) and Case Based Study (CBS), out of which DFA used for providing adaptive nature and shows the state transition according to their performance. Path of leaning varies with different student for a particular topic, this feature used to provide adaptively nature in E-learning. CBS used for providing study material based on the state. CBS uses case library to decide the study material.

**Literature 8**

**Identify-Based Distributed Provable Data Possession in Multicloud Storage**

**Author : Wang H**

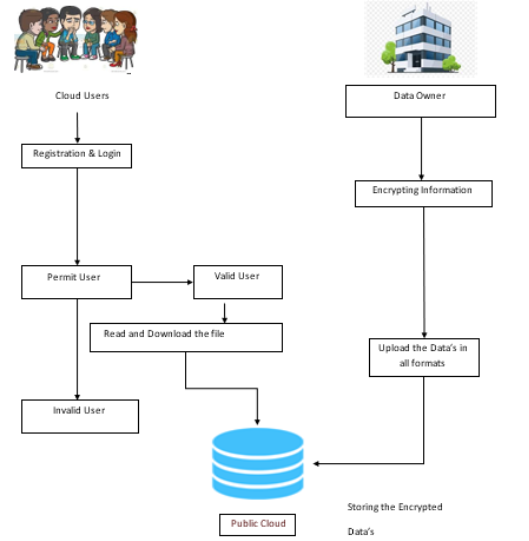
**LITERATURE DISCRIPTION :**

Remote data integrity checking is of crucial importance in cloud storage. It can make the clients verify whether their outsourced data is kept intact without downloading the whole data. In some application scenarios, the clients have to store their data on multicloud servers. At the same time, the integrity checking protocol must be efficient in order to save the verifier's cost. From the two points, we propose a novel remote data integrity checking model: ID-DPDP (identity-based distributed provable data possession) in multicloud storage. The formal system model and security model are given. Based on the bilinear pairings, a concrete ID-DPDP protocol is designed. The proposed ID-DPDP protocol is provably secure under the hardness assumption of the standard CDH (computational Diffie-Hellman) problem. In addition to the structural advantage of elimination of certificate management, our ID-DPDP protocol is also efficient and flexible. Based on the client's authorization, the proposed ID-DPDP protocol can realize private verification, delegated verification, and public verification.

1. **SYSTEM DESIGN**

**3.1 SYSTEM ARCHITECTURE:**

System architecture is the conceptual model that defines the structure, behaviour, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviours of the system. Figure shows the architecture.



**3.2 ARCHITECTURE DIAGRAM**

1. Browse and generate Digital sign ,Upload File by ABE(Attribute Based Enc)
2. View all uploaded files for corresponding owners
3. Delete file based on digital sign

CLOUD SERVER

1. View All Files
2. View Generated Secret Key
3. Give Permission For Dec Key
4. View All Transactions
5. View All Attackers
6. View Attackers Request
7. UnBlock The Attackers
8. View Results
9. View Time Delay Results
10. View Throughput Results

**Data Owner**

**1. View all corresponding owner keys.**

Res

**Finding Attackers**

Req

**KGC**

Request Secret Key

1. View all Data Owners and Authorize

2. View all End Users and Authorize

3. Generate Secret Key based on the file and send to end user

4. View all Generated Keys

**End User**

Register as doctor or professor or engineer

1. Request Secret Key
2. View Secret key response
3. Request Dec Key
4. View Dec key response
5. Search files based on keyword
6. Req Dec Key and view Response
7. Download file

ABE --- Attribute Based Encryption

**3.3 FLOW CHART**

KGC

Data owner

View all Data Owners and Authorize, View all End Users and Authorize

Register and login

Stores the files in cloud server

Generate Secret Key based on the file and send to end user

User view secret key response

View all Generated Keys, Provide attribute access permission

Cloud server verifies secret keys

Wrong

Correct

User

Download file from cloud server

**3.4 DATA FLOW DIAGRAM**

KGC

Data Owner

Response

End User

Cloud Server

Response

Request

Request

**3.5 DATA FLOW LEVELS**

**Level -0**

CLOUD SERVER

DATA OWNER

**Level-1**

KGC

End user

Cloud server

1. **MODULES**

The proposed system consists of four main modules. They are:

1. **END USER**

In this module, the user will register as a doctor or professor or engineer. The user will request for secret key and view the response, similarly the decrypt key and its response. The user will then search for files based on the keywords and download the corresponding files.

1. **CLOUD SERVER**

Cloud server will view all the uploaded files and give permission for the decrypt key and generates the secret key requested by the users for particular file. Cloud will see all the transactions related to files uploaded and the attackers and block the attackers and View Results,View Time Delay Results,View Throughput Results.

1. **DATA OWNER**

In this module, data owner will browse encrypt and upload the files with the digital sign in ABE (Attribute Based Encryption). Views all the uploaded files and delete based on digital sign.

1. **KGC**

In this module kgc will authorize both dataowner and end-users. KGC will generate the secret key based on the user’s requests and view all the generated keys and provide the attribute access permission.

1. **SYSTEM STUDY**

**5.1 PRELIMINARY INVESTIGATION**

The first and foremost strategy for development of a project starts from the thought of designing a mail enabled platform for a small firm in which it is easy and convenient of sending and receiving messages, there is a search engine ,address book and also including some entertaining games. When it is approved by the organization and our project guide the first activity, ie. preliminary investigation begins. The activity has three parts:

**Request Clarification**

**Feasibility Study**

**Request Approval**

**REQUEST CLARIFICATION**

After the approval of the request to the organization and project guide, with an investigation being considered, the project request must be examined to determine precisely what the system requires.

Here our project is basically meant for users within the company whose systems can be interconnected by the Local Area Network(LAN). In today’s busy schedule man need everything should be provided in a readymade manner. So taking into consideration of the vastly use of the net in day to day life, the corresponding development of the portal came into existence.

**5.2) FEASIBILITY ANALYSIS**

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resource and time. The different feasibilities that have to be analyzed are

* **Operational Feasibility**
* **Economic Feasibility**
* **Technical Feasibility**

###### **Operational Feasibility**

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

**Economic Feasibility**

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at anytime. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economically feasible.

###### **Technical Feasibility**

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform Independent environment. Java Server Pages, JavaScript, HTML, SQL server and WebLogic Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility.

**REQUEST APPROVAL**

Not all request projects are desirable or feasible. Some organization receives so many project requests from client users that only few of them are pursued. However, those projects that are both feasible and desirable should be put into schedule. After a project request is approved, it cost, priority, completion time and personnel requirement is estimated and used to determine where to add it to any project list. Truly speaking, the approval of those above factors, development works can be launched.

**5.3) INPUT DESIGN**

Input Design plays a vital role in the life cycle of software development, it requires very careful attention of developers. The input design is to feed data to the application as accurate as possible. So inputs are supposed to be designed effectively so that the errors occurring while feeding are minimized. According to Software Engineering Concepts, the input forms or screens are designed to provide to have a validation control over the input limit, range and other related validations.

This system has input screens in almost all the modules. Error messages are developed to alert the user whenever he commits some mistakes and guides him in the right way so that invalid entries are not made. Let us see deeply about this under module design.

Input design is the process of converting the user created input into a computer-based format. The goal of the input design is to make the data entry logical and free from errors. The error is in the input are controlled by the input design. The application has been developed in user-friendly manner. The forms have been designed in such a way during the processing the cursor is placed in the position where must be entered. The user is also provided with in an option to select an appropriate input from various alternatives related to the field in certain cases.

Validations are required for each data entered. Whenever a user enters an erroneous data, error message is displayed and the user can move on to the subsequent pages after completing all the entries in the current page.

5.4) OUTPUT DESIGN

The Output from the computer is required to mainly create an efficient method of communication within the company primarily among the project leader and his team members, in other words, the administrator and the clients. The output of VPN is the system which allows the project leader to manage his clients in terms of creating new clients and assigning new projects to them, maintaining a record of the project validity and providing folder level access to each client on the user side depending on the projects allotted to him. After completion of a project, a new project may be assigned to the client. User authentication procedures are maintained at the initial stages itself. A new user may be created by the administrator himself or a user can himself register as a new user but the task of assigning projects and validating a new user rests with the administrator only.

The application starts running when it is executed for the first time. The server has to be started and then the internet explorer in used as the browser. The project will run on the local area network so the server machine will serve as the administrator while the other connected systems can act as the clients. The developed system is highly user friendly and can be easily understood by anyone using it even for the first time.

1. **UML DIAGRAMS**

**6.1Use Case Diagram kgc**

: View all Data Owners and Authorize, View all End Users and Authorize

Browse and generate Digital sign

Generate Secret Key based on the file and send to end user

**Data Owner**

Upload File by ABE,View all uploaded files for corresponding owners

**Cloud Server**

Delete file based on digital sign

View all Files

View Generated Secret key

Give Permission for dec key

**End user**

View all transactions, View Results View Time Delay Results, View Throughput Results

Request Secret Key,View Secret key response

Request Dec Key,View Dec key response

Search files based on keyword, Req Dec Key and view Response, Download file.

**6.2 CLASS DIAGRAM**

DATA OWNER CLOUD SERVER

**METHODS**: View all Files,View Generated Secret key, Give Permission for dec key,View all transactions,View all attackers,Block the attacker, View Results ,View Time Delay Results ,View Throughput Results.

**MEMBERS:**

Cloud server name,password.

**METHODS:** Browse and generate Digital sign ,Upload File by ABE,View all uploaded files for corresponding owners,Delete file based on digital sign.

**MEMBERS:**

Owner name,password.

USER

KGC

**METHODS**: Request Secret Key,View Secret key response,Request Dec Key,View Dec key response,Search files based on keyword,Req Dec Key and view Response,Download file.

**MEMBERS:**

Name,password,email id,mobile number,date of birth,gender,address pincode,select image.

**Methods** : View all Data Owners and Authorize, View all End Users and Authorize,Generate Secret Key based on the file and send to end user, View all Generated Keys,

**MEMBERS :**

Authority name,password.

1. **SOFTWARE ENVIRONMENT**

## Java Technology

Java technology is both a programming language and a platform.

### The Java Programming Language

### The Java programming language is a high-level language that can be characterized by all of the following buzzwords:

* + - Simple
    - Architecture neutral
    - Object oriented
    - Portable
    - Distributed
    - High performance
    - Interpreted
    - Multithreaded
    - Robust
    - Dynamic
    - Secure

With most programming languages, you either compile or interpret a program so that you can run it on your computer. The Java programming language is unusual in that a program is both compiled and interpreted. With the compiler, first you translate a program into an intermediate language called Java byte codes —the platform-independent codes interpreted by the interpreter on the Java platform. The interpreter parses and runs each Java byte code instruction on the computer. Compilation happens just once; interpretation occurs each time the program is executed. The following figure illustrates how this works.



You can think of Java byte codes as the machine code instructions for the Java Virtual Machine (Java VM). Every Java interpreter, whether it’s a development tool or a Web browser that can run applets, is an implementation of the Java VM. Java byte codes help make “write once, run anywhere” possible. You can compile your program into byte codes on any platform that has a Java compiler. The byte codes can then be run on any implementation of the Java VM. That means that as long as a computer has a Java VM, the same program written in the Java programming language can run on Windows 2000, a Solaris workstation, or on an iMac.



### The Java Platform

A platform is the hardware or software environment in which a program runs. We’ve already mentioned some of the most popular platforms like Windows 2000, Linux, Solaris, and MacOS. Most platforms can be described as a combination of the operating system and hardware. The Java platform differs from most other platforms in that it’s a software-only platform that runs on top of other hardware-based platforms.

The Java platform has two components:

* The Java Virtual Machine (Java VM)
* The Java Application Programming Interface (Java API)

You’ve already been introduced to the Java VM. It’s the base for the Java platform and is ported onto various hardware-based platforms.

The Java API is a large collection of ready-made software components that provide many useful capabilities, such as graphical user interface (GUI) widgets. The Java API is grouped into libraries of related classes and interfaces; these libraries are known as packages. The next section, What Can Java Technology Do? Highlights what functionality some of the packages in the Java API provide.

The following figure depicts a program that’s running on the Java platform. As the figure shows, the Java API and the virtual machine insulate the program from the hardware.



Native code is code that after you compile it, the compiled code runs on a specific hardware platform. As a platform-independent environment, the Java platform can be a bit slower than native code. However, smart compilers, well-tuned interpreters, and just-in-time byte code compilers can bring performance close to that of native code without threatening portability.

## What Can Java Technology Do?

The most common types of programs written in the Java programming language are applets and applications. If you’ve surfed the Web, you’re probably already familiar with applets. An applet is a program that adheres to certain conventions that allow it to run within a Java-enabled browser.

However, the Java programming language is not just for writing cute, entertaining applets for the Web. The general-purpose, high-level Java programming language is also a powerful software platform. Using the generous API, you can write many types of programs.

An application is a standalone program that runs directly on the Java platform. A special kind of application known as a server serves and supports clients on a network. Examples of servers are Web servers, proxy servers, mail servers, and print servers. Another specialized program is a servlet. A servlet can almost be thought of as an applet that runs on the server side. Java Servlets are a popular choice for building interactive web applications, replacing the use of CGI scripts. Servlets are similar to applets in that they are runtime extensions of applications. Instead of working in browsers, though, servlets run within Java Web servers, configuring or tailoring the server.

How does the API support all these kinds of programs? It does so with packages of software components that provides a wide range of functionality. Every full implementation of the Java platform gives you the following features:

* **The essentials**: Objects, strings, threads, numbers, input and output, data structures, system properties, date and time, and so on.
* **Applets**: The set of conventions used by applets.
* **Networking**: URLs, TCP (Transmission Control Protocol), UDP (User Data gram Protocol) sockets, and IP (Internet Protocol) addresses.
* **Internationalization**: Help for writing programs that can be localized for users worldwide. Programs can automatically adapt to specific locales and be displayed in the appropriate language.
* **Security**: Both low level and high level, including electronic signatures, public and private key management, access control, and certificates.
* **Software components**: Known as JavaBeansTM, can plug into existing component architectures.
* **Object serialization**: Allows lightweight persistence and communication via Remote Method Invocation (RMI).
* **Java Database Connectivity (JDBCTM)**: Provides uniform access to a wide range of relational databases.

The Java platform also has APIs for 2D and 3D graphics, accessibility, servers, collaboration, telephony, speech, animation, and more. The following figure depicts what is included in the Java 2 SDK.



## How Will Java Technology Change My Life?

We can’t promise you fame, fortune, or even a job if you learn the Java programming language. Still, it is likely to make your programs better and requires less effort than other languages. We believe that Java technology will help you do the following:

* **Get started quickly**: Although the Java programming language is a powerful object-oriented language, it’s easy to learn, especially for programmers already familiar with C or C++.
* **Write less code**: Comparisons of program metrics (class counts, method counts, and so on) suggest that a program written in the Java programming language can be four times smaller than the same program in C++.
* **Write better code**: The Java programming language encourages good coding practices, and its garbage collection helps you avoid memory leaks. Its object orientation, its JavaBeans component architecture, and its wide-ranging, easily extendible API let you reuse other people’s tested code and introduce fewer bugs.
* **Develop programs more quickly**: Your development time may be as much as twice as fast versus writing the same program in C++. Why? You write fewer lines of code and it is a simpler programming language than C++.
* **Avoid platform dependencies with 100% Pure Java**: You can keep your program portable by avoiding the use of libraries written in other languages. The 100% Pure JavaTM Product Certification Program has a repository of historical process manuals, white papers, brochures, and similar materials online.
* **Write once, run anywhere**: Because 100% Pure Java programs are compiled into machine-independent byte codes, they run consistently on any Java platform.
* **Distribute software more easily**: You can upgrade applets easily from a central server. Applets take advantage of the feature of allowing new classes to be loaded “on the fly,” without recompiling the entire program.

### ODBC

Microsoft Open Database Connectivity (ODBC) is a standard programming interface for application developers and database systems providers. Before ODBC became a *de facto* standard for Windows programs to interface with database systems, programmers had to use proprietary languages for each database they wanted to connect to. Now, ODBC has made the choice of the database system almost irrelevant from a coding perspective, which is as it should be. Application developers have much more important things to worry about than the syntax that is needed to port their program from one database to another when business needs suddenly change.

Through the ODBC Administrator in Control Panel, you can specify the particular database that is associated with a data source that an ODBC application program is written to use. Think of an ODBC data source as a door with a name on it. Each door will lead you to a particular database. For example, the data source named Sales Figures might be a SQL Server database, whereas the Accounts Payable data source could refer to an Access database. The physical database referred to by a data source can reside anywhere on the LAN.

The ODBC system files are not installed on your system by Windows 95. Rather, they are installed when you setup a separate database application, such as SQL Server Client or Visual Basic 4.0. When the ODBC icon is installed in Control Panel, it uses a file called ODBCINST.DLL. It is also possible to administer your ODBC data sources through a stand-alone program called ODBCADM.EXE. There is a 16-bit and a 32-bit version of this program and each maintains a separate list of ODBC data sources.

From a programming perspective, the beauty of ODBC is that the application can be written to use the same set of function calls to interface with any data source, regardless of the database vendor. The source code of the application doesn’t change whether it talks to Oracle or SQL Server. We only mention these two as an example. There are ODBC drivers available for several dozen popular database systems. Even Excel spreadsheets and plain text files can be turned into data sources. The operating system uses the Registry information written by ODBC Administrator to determine which low-level ODBC drivers are needed to talk to the data source (such as the interface to Oracle or SQL Server). The loading of the ODBC drivers is transparent to the ODBC application program. In a client/server environment, the ODBC API even handles many of the network issues for the application programmer.

The advantages of this scheme are so numerous that you are probably thinking there must be some catch. The only disadvantage of ODBC is that it isn’t as efficient as talking directly to the native database interface. ODBC has had many detractors make the charge that it is too slow. Microsoft has always claimed that the critical factor in performance is the quality of the driver software that is used. In our humble opinion, this is true. The availability of good ODBC drivers has improved a great deal recently. And anyway, the criticism about performance is somewhat analogous to those who said that compilers would never match the speed of pure assembly language. Maybe not, but the compiler (or ODBC) gives you the opportunity to write cleaner programs, which means you finish sooner. Meanwhile, computers get faster every year.

**JDBC**

In an effort to set an independent database standard API for Java; Sun Microsystems developed Java Database Connectivity, or JDBC. JDBC offers a generic SQL database access mechanism that provides a consistent interface to a variety of RDBMSs. This consistent interface is achieved through the use of “plug-in” database connectivity modules, or *drivers*. If a database vendor wishes to have JDBC support, he or she must provide the driver for each platform that the database and Java run on.

To gain a wider acceptance of JDBC, Sun based JDBC’s framework on ODBC. As you discovered earlier in this chapter, ODBC has widespread support on a variety of platforms. Basing JDBC on ODBC will allow vendors to bring JDBC drivers to market much faster than developing a completely new connectivity solution.

JDBC was announced in March of 1996. It was released for a 90 day public review that ended June 8, 1996. Because of user input, the final JDBC v1.0 specification was released soon after.

The remainder of this section will cover enough information about JDBC for you to know what it is about and how to use it effectively. This is by no means a complete overview of JDBC. That would fill an entire book.

### JDBC Goals

Few software packages are designed without goals in mind. JDBC is one that, because of its many goals, drove the development of the API. These goals, in conjunction with early reviewer feedback, have finalized the JDBC class library into a solid framework for building database applications in Java.

The goals that were set for JDBC are important. They will give you some insight as to why certain classes and functionalities behave the way they do. The eight design goals for JDBC are as follows:

1. ***SQL Level API***

The designers felt that their main goal was to define a SQL interface for Java. Although not the lowest database interface level possible, it is at a low enough level for higher-level tools and APIs to be created. Conversely, it is at a high enough level for application programmers to use it confidently. Attaining this goal allows for future tool vendors to “generate” JDBC code and to hide many of JDBC’s complexities from the end user.

1. ***SQL Conformance***

SQL syntax varies as you move from database vendor to database vendor. In an effort to support a wide variety of vendors, JDBC will allow any query statement to be passed through it to the underlying database driver. This allows the connectivity module to handle non-standard functionality in a manner that is suitable for its users.

1. ***JDBC must be implemental on top of common database interfaces***The JDBC SQL API must “sit” on top of other common SQL level APIs. This goal allows JDBC to use existing ODBC level drivers by the use of a software interface. This interface would translate JDBC calls to ODBC and vice versa.
2. ***Provide a Java interface that is consistent with the rest of the Java system***

Because of Java’s acceptance in the user community thus far, the designers feel that they should not stray from the current design of the core Java system.

1. ***Keep it simple***

This goal probably appears in all software design goal listings. JDBC is no exception. Sun felt that the design of JDBC should be very simple, allowing for only one method of completing a task per mechanism. Allowing duplicate functionality only serves to confuse the users of the API.

1. ***Use strong, static typing wherever possible***

it adds to IP is a checksum for the contents of the datagram and port numbers. These are used to give a client/server model - see later.

**TCP**

TCP supplies logic to give a reliable connection-oriented protocol above IP. It provides a virtual circuit that two processes can use to communicate.

**Internet addresses**

In order to use a service, you must be able to find it. The Internet uses an address scheme for machines so that they can be located. The address is a 32 bit integer which gives the IP address. This encodes a network ID and more addressing. The network ID falls into various classes according to the size of the network address.

**Network address**

Class A uses 8 bits for the network address with 24 bits left over for other addressing. Class B uses 16 bit network addressing. Class C uses 24 bit network addressing and class D uses all 32.

### Subnet address

Internally, the UNIX network is divided into sub networks. Building 11 is currently on one sub network and uses 10-bit addressing, allowing 1024 different hosts.

**Host address**

8 bits are finally used for host addresses within our subnet. This places a limit of 256 machines that can be on the subnet.

### Total address



The 32 bit address is usually written as 4 integers separated by dots.

### Port addresses

A service exists on a host, and is identified by its port. This is a 16 bit number. To send a message to a server, you send it to the port for that service of the host that it is running on. This is not location transparency! Certain of these ports are "well known".

### Sockets

A socket is a data structure maintained by the system to handle network connections. A socket is created using the call socket. It returns an integer that is like a file descriptor. In fact, under Windows, this handle can be used with Read File and Write File functions.

#include <sys/types.h>

#include <sys/socket.h>

int socket(int family, int type, int protocol);

Here "family" will be AF\_INET for IP communications, protocol will be zero, and type will depend on whether TCP or UDP is used. Two processes wishing to communicate over a network create a socket each. These are similar to two ends of a pipe - but the actual pipe does not yet exist.

**JFree Chart**

JFreeChart is a free 100% Java chart library that makes it easy for developers to display professional quality charts in their applications. JFreeChart's extensive feature set includes:

A consistent and well-documented API, supporting a wide range of chart types;

A flexible design that is easy to extend, and targets both server-side and client-side applications;

Support for many output types, including Swing components, image files (including PNG and JPEG), and vector graphics file formats (including PDF, EPS and SVG);

JFreeChart is "open source" or, more specifically, [free software](http://www.gnu.org/philosophy/free-sw.html). It is distributed under the terms of the [GNU Lesser General Public Licence](http://www.gnu.org/licenses/lgpl.html) (LGPL), which permits use in proprietary applications.

## 1. Map Visualizations

Charts showing values that relate to geographical areas. Some examples include: (a) population density in each state of the United States, (b) income per capita for each country in Europe, (c) life expectancy in each country of the world. The tasks in this project include:

Sourcing freely redistributable vector outlines for the countries of the world, states/provinces in particular countries (USA in particular, but also other areas);

Creating an appropriate dataset interface (plus default implementation), a rendered, and integrating this with the existing XYPlot class in JFreeChart; Testing, documenting, testing some more, documenting some more.

## 2. Time Series Chart Interactivity

Implement a new (to JFreeChart) feature for interactive time series charts --- to display a separate control that shows a small version of ALL the time series data, with a sliding "view" rectangle that allows you to select the subset of the time series data to display in the main chart.

## 3. Dashboards

There is currently a lot of interest in dashboard displays. Create a flexible dashboard mechanism that supports a subset of JFreeChart chart types (dials, pies, thermometers, bars, and lines/time series) that can be delivered easily via both Java Web Start and an applet.

## 4. Property Editors

The property editor mechanism in JFreeChart only handles a small subset of the properties that can be set for charts. Extend (or reimplement) this mechanism to provide greater end-user control over the appearance of the charts.

**J2ME (Java 2 Micro edition):-**

Sun Microsystems defines J2ME as "a highly optimized Java run-time environment targeting a wide range of consumer products, including pagers, cellular phones, screen-phones, digital set-top boxes and car navigation systems." Announced in June 1999 at the JavaOne Developer Conference, J2ME brings the cross-platform functionality of the Java language to smaller devices, allowing mobile wireless devices to share applications. With J2ME, Sun has adapted the Java platform for consumer products that incorporate or are based on small computing devices.

**1. General J2ME architecture**



J2ME uses configurations and profiles to customize the Java Runtime Environment (JRE). As a complete JRE, J2ME is comprised of a configuration, which determines the JVM used, and a profile, which defines the application by adding domain-specific classes. The configuration defines the basic run-time environment as a set of core classes and a specific JVM that run on specific types of devices. We'll discuss configurations in detail in the The profile defines the application; specifically, it adds domain-specific classes to the J2ME configuration to define certain uses for devices. We'll cover profiles in depth in the The following graphic depicts the relationship between the different virtual machines, configurations, and profiles. It also draws a parallel with the J2SE API and its Java virtual machine. While the J2SE virtual machine is generally referred to as a JVM, the J2ME virtual machines, KVM and CVM, are subsets of JVM. Both KVM and CVM can be thought of as a kind of Java virtual machine -- it's just that they are shrunken versions of the J2SE JVM and are specific to J2ME.

**2.Developing J2ME applications**

Introduction In this section, we will go over some considerations you need to keep in mind when developing applications for smaller devices. We'll take a look at the way the compiler is invoked when using J2SE to compile J2ME applications. Finally, we'll explore packaging and deployment and the role preverification plays in this process.

**3.Design considerations for small devices**

Developing applications for small devices requires you to keep certain strategies in mind during the design phase. It is best to strategically design an application for a small device before you begin coding. Correcting the code because you failed to consider all of the "gotchas" before developing the application can be a painful process. Here are some design strategies to consider:

\* Keep it simple. Remove unnecessary features, possibly making those features a separate, secondary application.

\* Smaller is better. This consideration should be a "no brainer" for all developers. Smaller applications use less memory on the device and require shorter installation times. Consider packaging your Java applications as compressed Java Archive (jar) files.

\* Minimize run-time memory use. To minimize the amount of memory used at run time, use scalar types in place of object types. Also, do not depend on the garbage collector. You should manage the memory efficiently yourself by setting object references to null when you are finished with them. Another way to reduce run-time memory is to use lazy instantiation, only allocating objects on an as-needed basis. Other ways of reducing overall and peak memory use on small devices are to release resources quickly, reuse objects, and avoid exceptions.

**4.Configurations overview**

The configuration defines the basic run-time environment as a set of core classes and a specific JVM that run on specific types of devices. Currently, two configurations exist for J2ME, though others may be defined in the future:

\* **Connected Limited Device Configuration (CLDC)** is used specifically with the KVM for 16-bit or 32-bit devices with limited amounts of memory. This is the configuration (and the virtual machine) used for developing small J2ME applications. Its size limitations make CLDC more interesting and challenging (from a development point of view) than CDC. CLDC is also the configuration that we will use for developing our drawing tool application. An example of a small wireless device running small applications is a Palm hand-held computer.

\* **Connected Device Configuration (CDC)** is used with the C virtual machine (CVM) and is used for 32-bit architectures requiring more than 2 MB of memory. An example of such a device is a Net TV box.

**5.J2ME profiles**

**What is a J2ME profile?**

As we mentioned earlier in this tutorial, a profile defines the type of device supported. The Mobile Information Device Profile (MIDP), for example, defines classes for cellular phones. It adds domain-specific classes to the J2ME configuration to define uses for similar devices. Two profiles have been defined for J2ME and are built upon CLDC: KJava and MIDP. Both KJava and MIDP are associated with CLDC and smaller devices. Profiles are built on top of configurations. Because profiles are specific to the size of the device (amount of memory) on which an application runs, certain profiles are associated with certain configurations.

A skeleton profile upon which you can create your own profile, the Foundation Profile, is available for CDC.

**Profile 1: KJava**

KJava is Sun's proprietary profile and contains the KJava API. The KJava profile is built on top of the CLDC configuration. The KJava virtual machine, KVM, accepts the same byte codes and class file format as the classic J2SE virtual machine. KJava contains a Sun-specific API that runs on the Palm OS. The KJava API has a great deal in common with the J2SE Abstract Windowing Toolkit (AWT). However, because it is not a standard J2ME package, its main package is com.sun.kjava. We'll learn more about the KJava API later in this tutorial when we develop some sample applications.

**Profile 2: MIDP**

MIDP is geared toward mobile devices such as cellular phones and pagers. The MIDP, like KJava, is built upon CLDC and provides a standard run-time environment that allows new applications and services to be deployed dynamically on end user devices. MIDP is a common, industry-standard profile for mobile devices that is not dependent on a specific vendor. It is a complete and supported foundation for mobile application

development. MIDP contains the following packages, the first three of which are core CLDC packages, plus three MIDP-specific packages.

\* java.lang

\* java.io

\* java.util

\* javax.microedition.io

\* javax.microedition.lcdui

\* javax.microedition.midlet

\* javax.microedition.rms

#### **Client Server**

#### **Over view:**

With the varied topic in existence in the fields of computers, Client Server is one, which has generated more heat than light, and also more hype than reality. This technology has acquired a certain critical mass attention with its dedication conferences and magazines. Major computer vendors such as IBM and DEC, have declared that Client Servers is their main future market. A survey of DBMS magazine reveled that 76% of its readers were actively looking at the client server solution. The growth in the client server development tools from $200 million in 1992 to more than $1.2 billion in 1996.

Client server implementations are complex but the underlying concept is simple and powerful. A client is an application running with local resources but able to request the database and relate the services from separate remote server. The software mediating this client server interaction is often referred to as MIDDLEWARE.

The typical client either a PC or a Work Station connected through a network to a more powerful PC, Workstation, Midrange or Main Frames server usually capable of handling request from more than one client. However, with some configuration server may also act as client. A server may need to access other server in order to process the original client request.

The key client server idea is that client as user is essentially insulated from the physical location and formats of the data needs for their application. With the proper middleware, a client input from or report can transparently access and manipulate both local database on the client machine and remote databases on one or more servers. An added bonus is the client server opens the door to multi-vendor database access indulging heterogeneous table joins.

#### **What is a Client Server**

Two prominent systems in existence are client server and file server systems. It is essential to distinguish between client servers and file server systems. Both provide shared network access to data but the comparison dens there! The file server simply provides a remote disk drive that can be accessed by LAN applications on a file by file basis. The client server offers full relational database services such as SQL-Access, Record modifying, Insert, Delete with full relational integrity backup/ restore performance for high volume of transactions, etc. the client server middleware provides a flexible interface between client and server, who does what, when and to whom.

#### **Why Client Server**

Client server has evolved to solve a problem that has been around since the earliest days of computing: how best to distribute your computing, data generation and data storage resources in order to obtain efficient, cost effective departmental an enterprise wide data processing. During mainframe era choices were quite limited. A central machine housed both the CPU and DATA (cards, tapes, drums and later disks). Access to these resources was initially confined to batched runs that produced departmental reports at the appropriate intervals. A strong central information service department ruled the corporation. The role of the rest of the corporation limited to requesting new or more frequent reports and to provide hand written forms from which the central data banks were created and updated. The earliest client server solutions therefore could best be characterized as “SLAVE-MASTER”.

#### Time-sharing changed the picture. Remote terminal could view and even change the central data, subject to access permissions. And, as the central data banks evolved in to sophisticated relational database with non-programmer query languages, online users could formulate adhoc queries and produce local reports with out adding to the MIS applications software backlog. However remote access was through dumb terminals, and the client server remained subordinate to the Slave\Master.

#### Front end or User Interface Design

The entire user interface is planned to be developed in browser specific environment with a touch of Intranet-Based Architecture for achieving the Distributed Concept.

The browser specific components are designed by using the HTML standards, and the dynamism of the designed by concentrating on the constructs of the Java Server Pages.

#### **Communication or Database Connectivity Tier**

The Communication architecture is designed by concentrating on the Standards of Servlets and Enterprise Java Beans. The database connectivity is established by using the Java Data Base Connectivity.

The standards of three-tire architecture are given major concentration to keep the standards of higher cohesion and limited coupling for effectiveness of the operations.

#### **Features of The Language Used**

In my project, I have chosen *Java* language for developing the code.

#### **About Java**

Initially the language was called as “oak” but it was renamed as “Java” in 1995. The primary motivation of this language was the need for a platform-independent (i.e., architecture neutral) language that could be used to create software to be embedded in various consumer electronic devices.

* Java is a programmer’s language.
* Java is cohesive and consistent.
* Except for those constraints imposed by the Internet environment, Java gives the programmer, full control.

Finally, Java is to Internet programming where C was to system programming.

#### **Importance of Java to the Internet**

Java has had a profound effect on the Internet. This is because; Java expands the Universe of objects that can move about freely in Cyberspace. In a network, two categories of objects are transmitted between the Server and the Personal computer. They are: Passive information and Dynamic active programs. The Dynamic, Self-executing programs cause serious problems in the areas of Security and probability. But, Java addresses those concerns and by doing so, has opened the door to an exciting new form of program called the Applet.

#### **Java can be used to create two types of programs**

***Applications* and *Applets*:** An application is a program that runs on our Computer under the operating system of that computer. It is more or less like one creating using C or C++. Java’s ability to create Applets makes it important. An Applet is an application designed to be transmitted over the Internet and executed by a Java –compatible web browser. An applet is actually a tiny Java program, dynamically downloaded across the network, just like an image. But the difference is, it is an intelligent program, not just a media file. It can react to the user input and dynamically change.

#### **Features Of Java**

**Security**

Every time you that you download a “normal” program, you are risking a viral infection. Prior to Java, most users did not download executable programs frequently, and those who did scanned them for viruses prior to execution. Most users still worried about the possibility of infecting their systems with a virus. In addition, another type of malicious program exists that must be guarded against. This type of program can gather private information, such as credit card numbers, bank account balances, and passwords. Java answers both these concerns by providing a “firewall” between a network application and your computer.

When you use a Java-compatible Web browser, you can safely download Java applets without fear of virus infection or malicious intent.

**Portability**

For programs to be dynamically downloaded to all the various types of platforms connected to the Internet, some means of generating portable executable code is needed .As you will see, the same mechanism that helps ensure security also helps create portability. Indeed, Java’s solution to these two problems is both elegant and efficient.

**The Byte code**

The key that allows the Java to solve the security and portability problems is that the output of Java compiler is Byte code. Byte code is a highly optimized set of instructions designed to be executed by the Java run-time system, which is called the Java Virtual Machine (JVM). That is, in its standard form, the JVM is an interpreter for byte code.

Translating a Java program into byte code helps makes it much easier to run a program in a wide variety of environments. The reason is, once the run-time package exists for a given system, any Java program can run on it.

Although Java was designed for interpretation, there is technically nothing about Java that prevents on-the-fly compilation of byte code into native code. Sun has just completed its Just In Time (JIT) compiler for byte code. When the JIT compiler is a part of JVM, it compiles byte code into executable code in real time, on a piece-by-piece, demand basis. It is not possible to compile an entire Java program into executable code all at once, because Java performs various run-time checks that can be done only at run time. The JIT compiles code, as it is needed, during execution.

**Java, Virtual Machine (JVM)**

Beyond the language, there is the Java virtual machine. The Java virtual machine is an important element of the Java technology. The virtual machine can be embedded within a web browser or an operating system. Once a piece of Java code is loaded onto a machine, it is verified. As part of the loading process, a class loader is invoked and does byte code verification makes sure that the code that’s has been generated by the compiler will not corrupt the machine that it’s loaded on. Byte code verification takes place at the end of the compilation process to make sure that is all accurate and correct. So byte code verification is integral to the compiling and executing of Java code.

**Overall Description**

# Java Source

## Java byte code

# Java**VM**

Java

.Class

**Picture showing the development process of JAVA Program**

Java programming uses to produce byte codes and executes them. The first box indicates that the Java source code is located in a. Java file that is processed with a Java compiler called javac. The Java compiler produces a file called a. class file, which contains the byte code. The. Class file is then loaded across the network or loaded locally on your machine into the execution environment is the Java virtual machine, which interprets and executes the byte code.

**Java Architecture**

Java architecture provides a portable, robust, high performing environment for development. Java provides portability by compiling the byte codes for the Java Virtual Machine, which is then interpreted on each platform by the run-time environment. Java is a dynamic system, able to load code when needed from a machine in the same room or across the planet.

**Compilation of code**

When you compile the code, the Java compiler creates machine code (called byte code) for a hypothetical machine called Java Virtual Machine (JVM). The JVM is supposed to execute the byte code. The JVM is created for overcoming the issue of portability. The code is written and compiled for one machine and interpreted on all machines. This machine is called Java Virtual Machine.

**Compiling and interpreting Java Source Code**

During run-time the Java interpreter tricks the byte code file into thinking that it is running on a Java Virtual Machine. In reality this could be a Intel Pentium Windows 95 or Sun SARC station running Solaris or Apple Macintosh running system and all could receive code from any computer through Internet and run the Applets.

**Source**

**Code**

**………..**

**………..**

**………..**

**…………**

# PC Compiler

**Macintosh**

**Compiler**

**SPARC**

###### Compiler

**Java**

**Byte code**

**(Platform**

**indepen**

**dent)**

**Java**

**Interpreter**

**(PC)**

**Java**

**Interpreter**

**(Macintosh)**

**Java**

**Interpreter**

**(Sparc)**

**Simple**

Java was designed to be easy for the Professional programmer to learn and to use effectively. If you are an experienced C++ programmer, learning Java will be even easier. Because Java inherits the C/C++ syntax and many of the object oriented features of C++. Most of the confusing concepts from C++ are either left out of Java or implemented in a cleaner, more approachable manner. In Java there are a small number of clearly defined ways to accomplish a given task.

**Object-Oriented**

Java was not designed to be source-code compatible with any other language. This allowed the Java team the freedom to design with a blank slate. One outcome of this was a clean usable, pragmatic approach to objects. The object model in Java is simple and easy to extend, while simple types, such as integers, are kept as high-performance non-objects.

**Robust**

The multi-platform environment of the Web places extraordinary demands on a program, because the program must execute reliably in a variety of systems. The ability to create robust programs was given a high priority in the design of Java. Java is strictly typed language; it checks your code at compile time and run time.

Java virtually eliminates the problems of memory management and deallocation, which is completely automatic. In a well-written Java program, all run time errors can –and should –be managed by your program.

**JAVASCRIPT**

JavaScript is a script-based programming language that was developed by Netscape Communication Corporation. JavaScript was originally called Live Script and renamed as JavaScript to indicate its relationship with Java. JavaScript supports the development of both client and server components of Web-based applications. On the client side, it can be used to write programs that are executed by a Web browser within the context of a Web page. On the server side, it can be used to write Web server programs that can process information submitted by a Web browser and then updates the browser’s display accordingly.

Even though JavaScript supports both client and server Web programming, we prefer JavaScript at Client side programming since most of the browsers supports it. JavaScript is almost as easy to learn as HTML, and JavaScript statements can be included in HTML documents by enclosing the statements between a pair of scripting tags

<SCRIPTS>..</SCRIPT>.

<SCRIPT LANGUAGE = “JavaScript”>

JavaScript statements

</SCRIPT>

Here are a few things we can do with JavaScript :

* Validate the contents of a form and make calculations.
* Add scrolling or changing messages to the Browser’s status line.
* Animate images or rotate images that change when we move the mouse over them.
* Detect the browser in use and display different content for different browsers.
* Detect installed plug-ins and notify the user if a plug-in is required.

We can do much more with JavaScript, including creating entire application.

JavaScript Vs Java

JavaScript and Java are entirely different languages. A few of the most glaring differences are:

* Java applets are generally displayed in a box within the web document; JavaScript can affect any part of the Web document itself.
* While JavaScript is best suited to simple applications and adding interactive features to Web pages; Java can be used for incredibly complex applications.

There are many other differences but the important thing to remember is that JavaScript and Java are separate languages. They are both useful for different things; in fact they can be used together to combine their advantages.

**ADVANTAGES**

* JavaScript can be used for Sever-side and Client-side scripting.
* It is more flexible than VBScript.
* JavaScript is the default scripting languages at Client-side since all the browsers supports it.

Hyper Text Markup Language

Hypertext Markup Language (HTML), the languages of the World Wide Web (WWW), allows users to produces Web pages that include text, graphics and pointer to other Web pages (Hyperlinks).

HTML is not a programming language but it is an application of ISO Standard 8879, SGML (Standard Generalized Markup Language), but specialized to hypertext and adapted to the Web. The idea behind Hypertext is that instead of reading text in rigid linear structure, we can easily jump from one point to another point. We can navigate through the information based on our interest and preference. A markup language is simply a series of elements, each delimited with special characters that define how text or other items enclosed within the elements should be displayed. Hyperlinks are underlined or emphasized works that load to other documents or some portions of the same document.

HTML can be used to display any type of document on the host computer, which can be geographically at a different location. It is a versatile language and can be used on any platform or desktop.

HTML provides tags (special codes) to make the document look attractive. HTML tags are not case-sensitive. Using graphics, fonts, different sizes, color, etc., can enhance the presentation of the document. Anything that is not a tag is part of the document itself.

**Basic HTML Tags :**

**<!-- -->** Specifies comments

**<A>……….</A>** Creates hypertext links

**<B>……….</B>** Formats text as bold

**<BIG>……….</BIG>** Formats text in large font.

**<BODY>…</BODY>** Contains all tags and text in the HTML document

**<CENTER>...</CENTER>** Creates text

**<DD>…</DD>** Definition of a term

**<DL>...</DL>**  Creates definition list

**<FONT>…</FONT>** Formats text with a particular font

**<FORM>...</FORM>** Encloses a fill-out form

**<FRAME>...</FRAME>** Defines a particular frame in a set of frames

**<H#>…</H#>** Creates headings of different levels

**<HEAD>...</HEAD>** Contains tags that specify information about a document

**<HR>...</HR>** Creates a horizontal rule

**<HTML>…</HTML>** Contains all other HTML tags

**<META>...</META>** Provides meta-information about a document

**<SCRIPT>…</SCRIPT>** Contains client-side or server-side script

**<TABLE>…</TABLE>**  Creates a table

**<TD>…</TD>** Indicates table data in a table

**<TR>…</TR>** Designates a table row

**<TH>…</TH>** Creates a heading in a table

**ADVANTAGES**

* A HTML document is small and hence easy to send over the net. It is small because it does not include formatted information.
* HTML is platform independent.
* HTML tags are not case-sensitive.

**Java Database Connectivity**

**What Is JDBC?**

JDBC is a Java API for executing SQL statements. (As a point of interest, JDBC is a trademarked name and is not an acronym; nevertheless, JDBC is often thought of as standing for Java Database Connectivity. It consists of a set of classes and interfaces written in the Java programming language. JDBC provides a standard API for tool/database developers and makes it possible to write database applications using a pure Java API.

Using JDBC, it is easy to send SQL statements to virtually any relational database. One can write a single program using the JDBC API, and the program will be able to send SQL statements to the appropriate database. The combinations of Java and JDBC lets a programmer write it once and run it anywhere.

**What Does JDBC Do?**

Simply put, JDBC makes it possible to do three things:

* Establish a connection with a database
* Send SQL statements
* Process the results.

**JDBC versus ODBC and other APIs**

At this point, Microsoft's ODBC (Open Database Connectivity) API is that probably the most widely used programming interface for accessing relational databases. It offers the ability to connect to almost all databases on almost all platforms.

So why not just use ODBC from Java? The answer is that you can use ODBC from Java, but this is best done with the help of JDBC in the form of the JDBC-ODBC Bridge, which we will cover shortly. The question now becomes "Why do you need JDBC?" There are several answers to this question:

1. ODBC is not appropriate for direct use from Java because it uses a C interface. Calls from Java to native C code have a number of drawbacks in the security, implementation, robustness, and automatic portability of applications.
2. A literal translation of the ODBC C API into a Java API would not be desirable. For example, Java has no pointers, and ODBC makes copious use of them, including the notoriously error-prone generic pointer "void \*". You can think of JDBC as ODBC translated into an object-oriented interface that is natural for Java programmers.
3. ODBC is hard to learn. It mixes simple and advanced features together, and it has complex options even for simple queries. JDBC, on the other hand, was designed to keep simple things simple while allowing more advanced capabilities where required.
4. A Java API like JDBC is needed in order to enable a "pure Java" solution. When ODBC is used, the ODBC driver manager and drivers must be manually installed on every client machine. When the JDBC driver is written completely in Java, however, JDBC code is automatically installable, portable, and secure on all Java platforms from network computers to mainframes.

**Two-tier and Three-tier Models**

The JDBC API supports both two-tier and three-tier models for database access.

In the two-tier model, a Java applet or application talks directly to the database. This requires a JDBC driver that can communicate with the particular database management system being accessed. A user's SQL statements are delivered to the database, and the results of those statements are sent back to the user. The database may be located on another machine to which the user is connected via a network. This is referred to as a client/server configuration, with the user's machine as the client, and the machine housing the database as the server. The network can be an Intranet, which, for example, connects employees within a corporation, or it can be the Internet.

**JAVA**

**Application**

### JDBC

### DBMS

**Client machine**

**DBMS-proprietary protocol**

**Database server**

In the three-tier model, commands are sent to a "middle tier" of services, which then send SQL statements to the database. The database processes the SQL statements and sends the results back to the middle tier, which then sends them to the user. MIS directors find the three-tier model very attractive because the middle tier makes it possible to maintain control over access and the kinds of updates that can be made to corporate data. Another advantage is that when there is a middle tier, the user can employ an easy-to-use higher-level API which is translated by the middle tier into the appropriate low-level calls. Finally, in many cases the three-tier architecture can provide performance advantages.

**Java applet or**

**Html browser**

**Application**

**Server (Java)**

**JDBC**

## DBMS

**Client machine (GUI)**

**HTTP, RMI, or CORBA calls**

**Server machine (business Logic)**

**DBMS-proprietary protocol**

**Database server**

Until now the middle tier has typically been written in languages such as C or C++, which offer fast performance. However, with the introduction of optimizing compilers that translate Java byte code into efficient machine-specific code, it is becoming practical to implement the middle tier in Java. This is a big plus, making it possible to take advantage of Java's robustness, multithreading, and security features. JDBC is important to allow database access from a Java middle tier.

**JDBC Driver Types**

The JDBC drivers that we are aware of at this time fit into one of four categories:

* JDBC-ODBC bridge plus ODBC driver
* Native-API partly-Java driver
* JDBC-Net pure Java driver
* Native-protocol pure Java driver

**JDBC-ODBC Bridge**

If possible, use a Pure Java JDBC driver instead of the Bridge and an ODBC driver. This completely eliminates the client configuration required by ODBC. It also eliminates the potential that the Java VM could be corrupted by an error in the native code brought in by the Bridge (that is, the Bridge native library, the ODBC driver manager library, the ODBC driver library, and the database client library).

**What Is the JDBC- ODBC Bridge?**

The JDBC-ODBC Bridge is a JDBC driver, which implements JDBC operations by translating them into ODBC operations. To ODBC it appears as a normal application program. The Bridge implements JDBC for any database for which an ODBC driver is available. The Bridge is implemented as the

sun.jdbc.odbc Java package and contains a native library used to access ODBC. The Bridge is a joint development of Intersolv and JavaSoft.

**Java Server Pages (JSP)**

Java server Pages is a simple, yet powerful technology for creating and maintaining dynamic-content web pages. Based on the Java programming language, Java Server Pages offers proven portability, open standards, and a mature re-usable component model .The Java Server Pages architecture enables the separation of content generation from content presentation. This separation not eases maintenance headaches, it also allows web team members to focus on their areas of expertise. Now, web page designer can concentrate on layout, and web application designers on programming, with minimal concern about impacting each other’s work.

**Features of JSP**

**Portability:**

Java Server Pages files can be run on any web server or web-enabled application server that provides support for them. Dubbed the JSP engine, this support involves recognition, translation, and management of the Java Server Page lifecycle and its interaction components.

**Components**

It was mentioned earlier that the Java Server Pages architecture can include reusable Java components. The architecture also allows for the embedding of a scripting language directly into the Java Server Pages file. The components current supported include Java Beans, and Servlets.

**Processing**

A Java Server Pages file is essentially an HTML document with JSP scripting or tags. The Java Server Pages file has a JSP extension to the server as a Java Server Pages file. Before the page is served, the Java Server Pages syntax is parsed and processed into a Servlet on the server side. The Servlet that is generated outputs real content in straight HTML for responding to the client.

**Access Models:**

A Java Server Pages file may be accessed in at least two different ways. A client’s request comes directly into a Java Server Page. In this scenario, suppose the page accesses reusable Java Bean components that perform particular well-defined computations like accessing a database. The result of the Beans computations, called result sets is stored within the Bean as properties. The page uses such Beans to generate dynamic content and present it back to the client.

In both of the above cases, the page could also contain any valid Java code. Java Server Pages architecture encourages separation of content from presentation.

**Steps in the execution of a JSP Application:**

1. The client sends a request to the web server for a JSP file by giving the name of the JSP file within the form tag of a HTML page.
2. This request is transferred to the JavaWebServer. At the server side JavaWebServer receives the request and if it is a request for a jsp file server gives this request to the JSP engine.
3. JSP engine is program which can understands the tags of the jsp and then it converts those tags into a Servlet program and it is stored at the server side. This Servlet is loaded in the memory and then it is executed and the result is given back to the JavaWebServer and then it is transferred back to the result is given back to the JavaWebServer and then it is transferred back to the client.

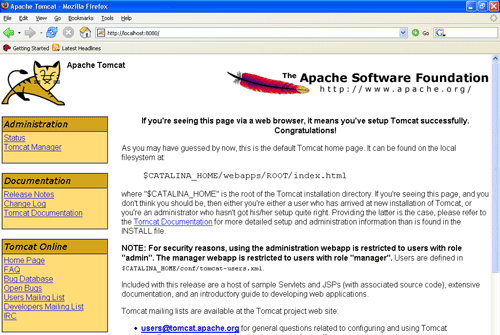
**JDBC connectivity**

The JDBC provides database-independent connectivity between the J2EE platform and a wide range of tabular data sources. JDBC technology allows an Application Component Provider to:

* Perform connection and authentication to a database server
* Manager transactions
* Move SQL statements to a database engine for preprocessing and execution
* Execute stored procedures
* Inspect and modify the results from Select statements.

# **Tomcat 6.0 web server**

Tomcat is an open source web server developed by Apache Group. Apache Tomcat is the servlet container that is used in the official Reference Implementation for the Java Servlet and Java Server Pages technologies. The Java Servlet and Java Server Pages specifications are developed by Sun under the Java Community Process. Web Servers like Apache Tomcat support only web components while an application server supports web components as well as business components (BEAs Weblogic, is one of the popular application server).To develop a web application with jsp/servlet install any web server like JRun, Tomcat etc to run your application.



1. **SYSTEM TESTING**

### TESTING METHODOLOGIES

The following are the Testing Methodologies:

* **Unit Testing.**
* **Integration Testing.**
* **User Acceptance Testing.**
* **Output Testing.**
* **Validation Testing.**

**8.1 Unit Testing**

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module’s control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All important processing path are tested for the expected results. All error handling paths are also tested.

**8.2 Integration Testing**

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

**The following are the types of Integration Testing:**

1. **Top Down Integration**

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

**2. Bottom-up Integration**

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

* The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
* A driver (i.e.) the control program for testing is written to coordinate test case input and output.
* The cluster is tested.
* Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

**8.3 User Acceptance Testing**

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

**8.4 Output Testing**

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

**8.5 Validation Checking**

Validation checks are performed on the following fields.

**Text Field:**

The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables. Incorrect entry always flashes and error message.

**Numeric Field:**

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to perform. Each module is subjected to test run along with sample data. The individually tested modules are integrated into a single system. Testing involves executing the real data information is used in the program the existence of any program defect is inferred from the output. The testing should be planned so that all the requirements are individually tested.

A successful test is one that gives out the defects for the inappropriate data and produces and output revealing the errors in the system.

**Preparation of Test Data**

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using that test data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

**Using Live Test Data:**

Live test data are those that are actually extracted from organization files. After a system is partially constructed, programmers or analysts often ask users to key in a set of data from their normal activities. Then, the systems person uses this data as a way to partially test the system. In other instances, programmers or analysts extract a set of live data from the files and have them entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And, although it is realistic data that will show how the system will perform for the typical processing requirement, assuming that the live data entered are in fact typical, such data generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

**Using Artificial Test Data:**

Artificial test data are created solely for test purposes, since they can be generated to test all combinations of formats and values. In other words, the artificial data, which can quickly be prepared by a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrote the programs. Often, an independent team of testers formulates a testing plan, using the systems specifications

The package “Virtual Private Network” has satisfied all the requirements specified as per software requirement specification and was accepted.

1. **CONCLUSION**

In this paper, we introduce a large universe searchable encryption scheme to protect the security of cloud storage system, which realizes regular language encryption and DFA search function. The cloud service provider could test whether the encrypted regular language in the encrypted ciphertext is acceptable by the DFA embedded in the submitted search token. In the test procedure, no plaintext of the regular language or the DFA will be leaked to the cloud server. We also put forth a concrete construction with lightweight encryption and token generation algorithms. An example is given to show how the system works. The proposed scheme is privacy-preserving and indistinguishable against KGA, which are proved in standard model. The comparison and experiment result confirm the low transmission and computation overhead of the scheme.

1. **Bibliography:**

[1] Erl T, Cope R, Naserpour A. Cloud computing design patterns[M].

Prentice Hall Press, 2015.

[2] Li Z, Dai Y, Chen G, et al. Toward network-level efficiency for cloud

storage services[M]//Content Distribution for Mobile Internet: A

Cloud-based Approach. Springer Singapore, 2016: 167-196.

[3] Sookhak M, Gani A, KhanMK, et al. Dynamic remote data auditing

for securing big data storage in cloud computing[J]. Information

Sciences, 2017, 380: 101-116.

[4] Zhang Q, Yang L T, Chen Z, Li P. Privacy-preserving doubleprojection deep computation model with crowdsourcing on cloud for big data feature learning[J]. IEEE Internet of Things Journal, 2017, DOI: 10.1109/JIOT.2017.2732735.

[5] Zhang Q, Yang L T, Chen Z, Li P. PPHOPCM: Privacy-preserving High-order Possibilistic c-Means Algorithm for Big Data Clustering with Cloud Computing[J]. IEEE Transactions on Big Data, 2017,DOI: 10.1109/TBDATA.2017.2701816.

[6] Liu J K, Liang K, Susilo W, et al. Two-factor data security protection mechanism for cloud storage system[J]. IEEE Transactions on Computers, 2016, 65(6): 1992-2004.

[7] Boneh D, Waters B. Conjunctive, subset, and range queries on encrypted data[C]//Theory of Cryptography Conference. Springer Berlin Heidelberg, 2007: 535-554.

[8] Q. Zheng, S. Xu, and G. Ateniese. VABKS: verifiable attribute-based keyword search over outsourced encrypted data. In INFOCOM, pp.522C530. IEEE, 2014.

[9] Liang K, Huang X, Guo F, et al. Privacy-Preserving and Regular Language Search Over Encrypted Cloud Data[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(10): 2365-2376.

[10] Chang V, Ramachandran M. Towards achieving data security with the cloud computing adoption framework [J]. IEEE Transactions on Services Computing, 2016, 9(1): 138-151.

[11] Zheng X H, Chen N, Chen Z, et al. Mobile cloud based framework for remote-resident multimedia discovery and access[J]. Journal of Internet Technology, 2014, 15(6): 1043-1050.

[12] Chang V, Kuo Y H, Ramachandran M. Cloud computing adoption framework: A security framework for business clouds [J]. Future Generation Computer Systems, 2016, 57: 24-41.

[13] Barsoum A. Provable data possession in single cloud server: A survey, classification and comparative study[J]. International Journal of Computer Applications, 2015, 123(9).

[14] Wang H. Identity-based distributed provable data possession in multicloud storage[J]. IEEE Transactions on Services Computing,2015, 8(2): 328-340.

[15] J, Tan X, Chen X, et al. Opor: Enabling proof of retrievability in cloud computing with resource-constrained devices[J]. IEEE Transactions on cloud computing, 2015, 3(2): 195-205.

[16] Tiwari D, Gangadharan G R. A novel secure cloud storage architecture combining proof of retrievability and revocation[C]//Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on. IEEE, 2015: 438-445.

[17] Omote K, Thao T P. MD-POR: multisource and direct repair for network coding-based proof of retrievability[J]. International Journal of Distributed Sensor Networks, 2015, 2015: 3.

[18] Hopcroft JE, Motwani R, Ullman JD. Automata theory, languages, and computation. International Edition 24 (2006).