Efficient and Expressive Keyword Search over Encrypted Data in Cloud

Tatimakula Rajasekhar[1], Smt. K.Vijaya Lakshmi[2]
[1,2]*M.C.A, Ph.D., Sri Venkateswara University, Tirupathi*

Abstract- **In this project Searchable encryption allows a cloud server to conduct keyword search over encrypted data on behalf of the data users without learning the underlying plaintexts. However, most existing searchable encryption schemes only support single or conjunctive keyword search, while a few other schemes that are able to perform expressive keyword search are computationally inefficient since they are built from bilinear pairings over the composite-order groups. In this paper, we propose an expressive public-key searchable encryption scheme in the prime-order groups, which allows keyword search policies (i.e., predicates, access structures) to be expressed in conjunctive, disjunctive or any monotonic Boolean formulas and achieves significant performance improvement over existing schemes. We formally define its security, and prove that it is selectively secure in the standard model. Also, we implement the proposed scheme using a rapid prototyping tool called Charm and conduct several experiments to eva ent than the ones built over the composite-order groups . Keyword research is one luate it performance. The results demonstrate that our scheme is much more effici of the most important, valuable, and high return activities in the search marketing field. Ranking for the right keywords can make or break your website. By researching your market's keyword demand, you can not only learn which terms and phrases to target with SEO, but also learn more about your customers as a whole.**

EXISTING SYSTEMS

However, the solution in as well as other existing PEKS schemes which improve on only support equality queries As such, our scheme is not only capable of expressive multi-keyword search, but also significantly more efficient than existing schemes built in composite-order groups. Using a randomness splitting technique, our scheme achieves security against offline keyword dictionary guessing attacks to the ciphertexts. Moreover, to preserve the privacy of keywords against offline keyword dictionary guessing attacks to trapdoors, we divide each keyword into keyword name and keyword value and assign a designated cloud server to conduct search operations in our construction.We formalize the security definition of expressive SE, and formally prove that our proposed expressive SE scheme is selectively secure in the standard model.

Disadvantages:

The approach based on set intersection leaks extra information to the cloud server beyond the results of the conjunctive query, whilst the approach using meta keywords require 2m meta keywords to accommodate all the possible conjunctive queries form keywords. It is straightforward to see that compared to the existing ones, our construction make a good balance in that it allows unbounded keywords, supports expressive access structures, and is built in the prime-order groups

PROPOSED SYSTEMS

We propose an expressive public-key searchable encryption scheme in the prime-order groups, which allows keyword search policies (i.e., predicates, access structures) to be expressed in conjunctive, disjunctive or any monotonic Boolean formulas and achieves significant performance improvement over existing schemes. We formally define its security, and prove that it is selectively secure in the standard model.In order to tackle the keyword search problem in the cloud-based healthcare information system scenario, we resort to public-key encryption with keyword search (PEKS) schemes, which is firstly proposed in . In a PEKS scheme, a ciphertext of the keywords called “PEKS ciphertext” is appended to an encrypted PHR. To retrieve all the encrypted PHRs containing a keyword, say “Diabetes”, a user sends a “trapdoor” associated with a search query on the keyword “Diabetes” to the cloud service provider, which selects all the encrypted PHRs containing the

keyword "Diabetes" and returns them to the user while without learning the underlying PHRs.

ADVANTAGES

The expressive SE scheme inherits the advantages of the Rouselakis-Waters scheme. Thus, it is straightforward to see that in our SE scheme, the size of the public parameter is immutable with the number of keywords, and the number of the keywords allowed for the system is unlimited and can be freely set. According to the analysis in terms of the pairing-friendly elliptic curves, prime order groups have a clear advantage in the parameter sizes over composite order groups. the advantage for a polynomial time adversary that can distinguish between the games Game0 and Game1 is negligible.

IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

Modules:

In this project we have following Four modules

- Searchable Encryption,
- Cloud Computing,
- Expressiveness Keyword search
- Attribute-Based Encryption

Cloud Computing:-

Cloud computing is a type of Internet-based computing that provides shared computer processing resources and data to computers and other devices on demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources (e.g., computer networks, servers, storage, applications and services) which can be rapidly provisioned and released with minimal management effort. Cloud-based healthcare information system that hosts outsourced personal health records (PHRs) from various healthcare providers. The PHRs are encrypted in order to comply with privacy regulations like HIPAA. In order to facilitate data use and sharing, it is highly desirable to have a searchable encryption (SE) scheme which allows the cloud service provider to search over encrypted PHRs on behalf of the authorized users (such as medical researchers or doctors) without learning information about the underlying plaintext. Note that the context we are considering supports private data sharing among multiple data providers and multiple data users.

Attribute-based encryption:-

Attribute-based encryption is a type of public-key encryption in which the secret key of a user and the ciphertext are dependent upon attributes (e.g. the country in which he lives, or the kind of subscription he has). The basic idea of our scheme is to modify a key-policy attributed-based encryption (KP-ABE) scheme constructed from bilinear pairing over prime-order groups. Without loss of generality, we will use the large universe KP-ABE scheme selectively secure in the standard model proposed by Rouselakis and Waters in [to illustrate our construction during the rest of the paper.

Expressiveness Keyword:-

The proposed scheme should support keyword access structures expressed in any Boolean formula with AND and OR gates. Efficiency. The proposed scheme should be adequately efficient in terms of computation, communication and storage for practical applications.

A ciphertext without its corresponding trapdoors should not disclose any information about the keyword values it contains to the cloud server and outsiders. Second, a trapdoor should not leak information on keyword values to any outside attackers without the private key of the designated cloud server.

We capture this notion of security for the SE scheme in terms of semantic security to ensure that encrypted data does not reveal any information about the keyword values, which we call "selective in distinguishability against chosen keyword-set .

System Configuration:

HARDWARE REQUIREMENTS:

Hardware	- Pentium
Speed	- 1.1 GHz
RAM	- 1GB
Hard Disk	- 20 GB
Floppy Drive	- 1.44 MB
Key Board	- Standard Windows Keyboard
Mouse	- Two or Three Button Mouse
Monitor	- SVGA

SOFTWARE REQUIREMENTS:

Operating System	: Windows
Technology	: Java and J2EE
Web Technologies	: Html, JavaScript, CSS
IDE	: My Eclipse
Web Server	: Tomcat
Tool kit	: Android Phone
Database	: My SQL
Java Version	: J2SDK1.5

System Configuration:

HARDWARE REQUIREMENTS:

Hardware	- Pentium
Speed	- 1.1 GHz
RAM	- 1GB
Hard Disk	- 20 GB
Floppy Drive	- 1.44 MB
Key Board	- Standard Windows Keyboard
Mouse	- Two or Three Button Mouse
Monitor	- SVGA

SOFTWARE REQUIREMENTS:

Operating System	: Windows
Technology	: Java and J2EE
Web Technologies	: Html, JavaScript, CSS
IDE	: My Eclipse
Web Server	: Tomcat
Tool kit	: Android Phone
Database	: My SQL
Java Version	: J2SDK1.5

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for

immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- ❖ Convey information about past activities, current status or projections of the
- ❖ Future.
- ❖ Signal important events, opportunities, problems, or warnings.
- ❖ Trigger an action.
- ❖ Confirm an action.

STUDY OF THE SYSTEM

To provide flexibility to the users, the interfaces have been developed that are accessible through a browser. The GUI'S at the top level have been categorized as

1. Administrative user interface
2. The operational or generic user interface

The 'administrative user interface' concentrates on the consistent information that is practically, part of the organizational activities and which needs proper authentication for the data collection. These interfaces help the administrators with all the transactional states like Data insertion, Data deletion and Date updation along with the extensive data search capabilities.

The 'operational or generic user interface' helps the end users of the system in transactions through the existing data and required services. The operational user interface also helps the ordinary users in managing their own information in a customized manner as per the included flexibilities.

INPUT & OUTPOUT REPRESENTETION

Input design is a part of overall system design. The main objective during the input design is as given below:

- To produce a cost-effective method of input.
- To achieve the highest possible level of accuracy.
- To ensure that the input is acceptable and understood by the user.

INPUT STAGES:

The main input stages can be listed as below:

- Data recording
- Data transcription
- Data conversion
- Data verification
- Data control
- Data transmission
- Data validation
- Data correction

INPUT TYPES

It is necessary to determine the various types of inputs. Inputs can be categorized as follows:

- External inputs, which are prime inputs for the system.
- Internal inputs, which are user communications with the system.
- Operational, which are computer department's communications to the system?
- Interactive, which are inputs entered during a dialogue.

INPUT MEDIA

At this stage choice has to be made about the input media. To conclude about the input media consideration has to be given to;

- Type of input
- Flexibility of format
- Speed
- Accuracy
- Verification methods
- Rejection rates
- Ease of correction
- Storage and handling requirements
- Security

- Easy to use
- Portability

Keeping in view the above description of the input types and input media, it can be said that most of the inputs are of the form of internal and interactive. As Input data is to be the directly keyed in by the user, the keyboard can be considered to be the most suitable input device.

OUTPUT DESIGN

In general are:

- External Outputs whose destination is outside the organization.
- Internal Outputs whose destination is with in organization and they are the User's main interface with the computer. Outputs from computer systems are required primarily to communicate the results of processing to users. They are also used to provide a permanent copy of the results for later consultation. The various types of outputs
- Operational outputs whose use is purely with in the computer department.
- Interface outputs, which involve the user in communicating directly with the system.

MANAGEMENT

Achieving high-speed development is a complex process. Systems will not be developed and deployed rapidly if bureaucracy and political obstacles stand in the way or if users are not appropriately involved. Management must be totally committed to RAD in order to manage the change in culture. They must be prepared to motivate both users and IT staff, select and manage SWAT teams, and demonstrate through the use of performance measurements that RAD does mean speed, quality, and productivity. Good management and dedication to the ideals of Rapid Application Development are thus essential to faster system building.

To successfully introduce rapid development, management must pay careful attention to human motivation. Managers should target those professionals whom they deem as 'Early Adapters.' 'Early Adapters' are those people who see the value of a new methodology and lead the way in making it practical to use. These employees are enthusiastic about the new methodology and they want to make it work well in their environment. Similarly, managers must be aware of the type of motivation that is most effective for each individual employee, whether it be money, pride, prestige, excitement, or some combination thereof.

Because Rapid Application Development is such a sweeping change from the conventional development methods, the best way for a manager to introduce new rapid development techniques is to start small. Original Construction Teams of two to four people should be established and their members should be thoroughly trained in the use of the tools and techniques. As these teams gain experience, they will be able to fine-tune the development lifecycle to improve its effectiveness in their environment. Underlying all of this progress, however, managers must remember the importance of comprehensive and quality training in the use of tools. Good training with tools that are exciting to use can have a profound impact on the attitude of IT professionals, as well as ensure the uninterrupted success of the rapid development project.

The RAD modelis a linear sequential software development process that emphasizes an extremely short development cycle. The RAD model is a "high speed" adaptation of the linear sequential model in which rapid development is achieved by using a component-based construction approach. Used primarily for information systems applications, the RAD approach encompasses the following phases:

1. Business modeling

The information flow among business functions is modeled in a way that answers the following questions:

What information drives the business process?
What information is generated?
Who generates it?
Where does the information go?
Who processes it?

2. Data modeling

The information flow defined as part of the business modeling phase is refined into a set of data objects that are needed to support the business. The characteristic (called attributes) of each object is

identified and the relationships between these objects are defined.

3. Process modeling

The data objects defined in the data-modeling phase are transformed to achieve the information flow necessary to implement a business function. Processing the descriptions are created for adding, modifying, deleting, or retrieving a data object.

4. Application generation

The RAD model assumes the use of the RAD tools like VB, VC++, Delphi etc... rather than creating software using conventional third generation programming languages. The RAD model works to reuse existing program components (when possible) or create reusable components (when necessary). In all cases, automated tools are used to facilitate construction of the software.

5. Testing and turnover

Since the RAD process emphasizes reuse, many of the program components have already been tested. This minimizes the testing and development time.

SDLC is nothing but Software Development Life Cycle. It is a standard which is used by software industry to develop good software.

Stages in SDLC:

- Requirement Gathering
- Analysis
- Designing
- Coding
- Testing
- Maintenance

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

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.

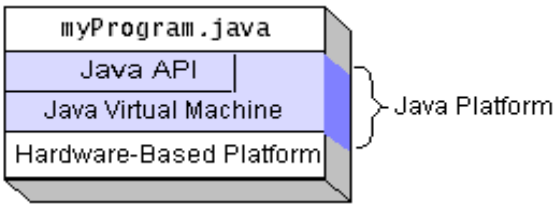
The Java platform has two components:

- 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.

Cloud Computing:-

Cloud computing is a type of Internet-based computing that provides shared computer processing resources and data to computers and other devices on demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources (e.g., computer networks, servers, storage, applications and services) which can be rapidly provisioned and released with minimal management effort. Cloud-based healthcare information system that hosts outsourced personal health records (PHRs) from various healthcare providers. The PHRs are encrypted in order to comply with privacy regulations like HIPAA. In order to facilitate data use and sharing, it is highly desirable to have a searchable encryption (SE) scheme which allows the cloud service provider to search over encrypted PHRs on behalf of the authorized users (such as medical researchers or doctors) without learning information about the underlying plaintext. Note that the context we are considering supports private data sharing among multiple data providers and multiple data users.

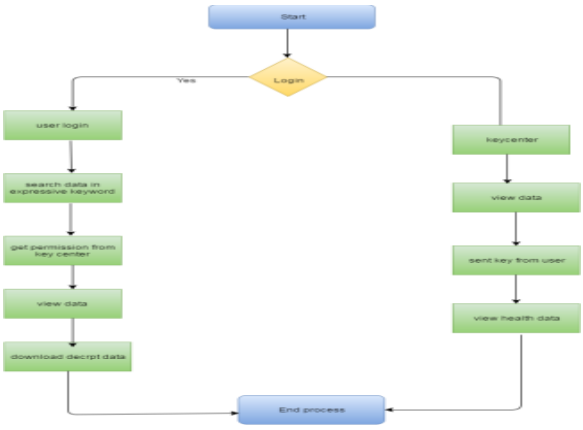
Data Flow Diagram / Use Case Diagram / Flow Diagram

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of the input data to the system, various processing carried out on these data, and the output data is generated by the system.

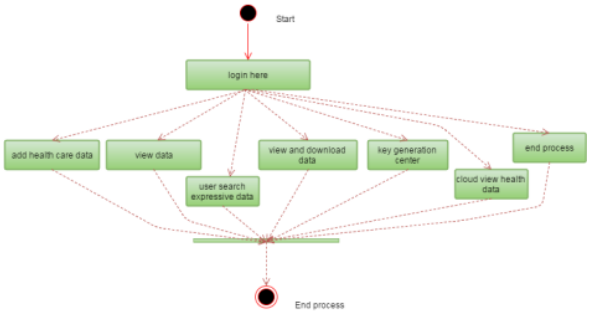
Flow Diagram:

User and Trapdoor key center:-

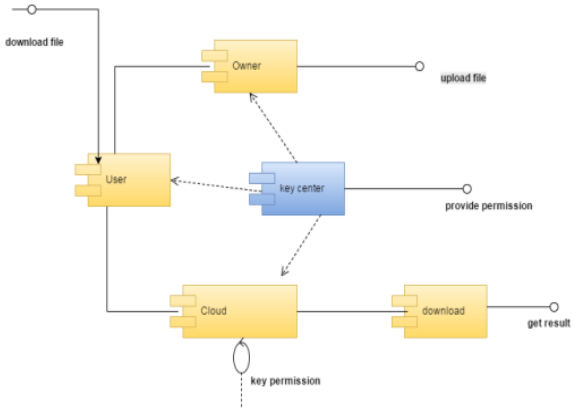
Owner And Cloud server:-



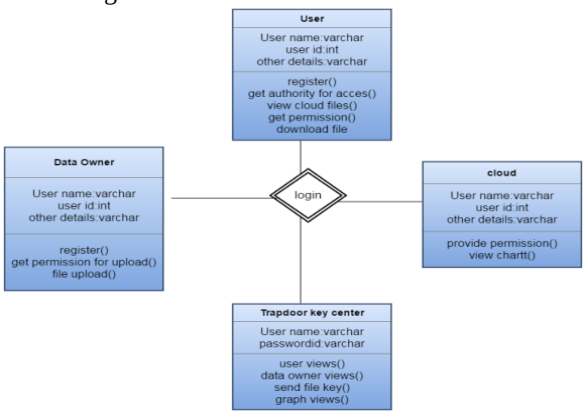
Activity Diagrams:-



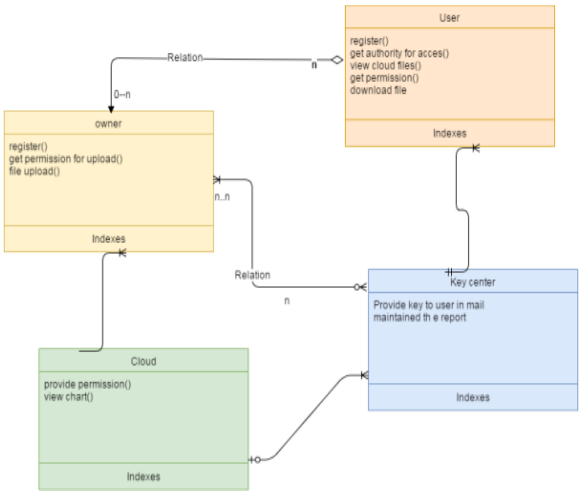
Component Diagrams:-



Class Diagrams:-



ER-Diagrams:-



CONCLUSION

In order to allow a cloud server to search on encrypted data without learning the underlying plaintexts in the public key setting, Boneh proposed a cryptographic primitive called public-key encryption with keyword search (PEKS). Since then, considering different requirements in practice, e.g., communication overhead, searching criteria and security enhancement, various kinds of searchable encryption systems have been put forth. However, there exist only a few public-key searchable encryption systems that support expressive keyword search policies, and they are all built from the inefficient composite-order groups . In this paper, we focused on the design and analysis of public-key sive searching formulas. Based on a large universe key-policy attribute-based encryption scheme given in , we presented an expressive searchable encryption system in the prime order group which supports expressive access structures expressed in any monotonic Boolean formulas. Also, we proved its security in the standard model, and analyzed its efficiency using computer simulations.

REFERENCES

[1] R. Bekkerman, R. El-Yaniv, N. Tishby, and Y.Winter. Distributional word clusters vs. words for text categorization. Journal of Machine Learning Research, 3:1183–1208, 2003.

[2] J. Bi, K. P. Bennett,M. J. Embrechts, C. M. Breneman, and M. Song. Dimensionality reduction via sparse support vector machines. Journal of Machine Learning Research, 3:1229–1243, 2003.

[3] G. Cavallanti, N. Cesa-Bianchi, and C. Gentile. Tracking the best hyperplane with a simple budget perceptron. Machine Learning, 69(2-3):143–167, 2007.

[4] N. Cesa-Bianchi, S. Shalev-Shwartz, and O. Shamir. Efficient learning with partially observed attributes. Journal of Machine Learning Research, pages 2857–2878, 2011.

[5] A. B. Chan, N. Vasconcelos, and G. R. G. Lanckriet. Direct convex relaxations of sparse svm. In ICML, pages 145–153, 2007.

[6] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer. Online passive-aggressive algorithms. J. Mach. Learn. Res. (JMLR), 7:551–585, 2006.

[7] K. Crammer, M. Dredze, and F. Pereira. Exact convex confidenceweighted learning. In NIPS, pages 345–352, 2008.

[8] K. Crammer, A. Kulesza, and M. Dredze. Adaptive regularization of weight vectors. In NIPS, pages 414–422, 2009.

[9] M. Dash and V. Gopalkrishnan. Distance based feature selection for clustering microarray data. In DASFAA, pages 512–519, 2008.

[10] M. Dash and H. Liu. Feature selection for classification. Intell. Data Anal., 1(1-4):131–156, 1997.

[11] O. Dekel, S. Shalev-Shwartz, and Y. Singer. The forgetron: A kernel-based perceptron on a budget. SIAM J. Comput., 37(5):1342–1372, 2008.

[12] C. H. Q. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. J. Bioinformatics and Computational Biology, 3(2):185–206, 2005.