Fast Phrase Search for Encrypted Cloud Storage

CSCI 5900

TEAM PROJECT

ADVANCED MOBILITY AND CLOUD COMPUTING

Contents

1.Abstract:

Cloud Computing, the current trend in terms of data storage and data retrieval remotely. When it comes to challenges in cloud computing same thing that is, storage and retrieval are on the top priority. There has been a lot of research on searching data in cloud environment, but mainly most of the research has been concentrated on conjunctive keyword search which becomes less accuracy as the data size increases and inclusion-relation attacks will be more frequent. Less efforts have been done on specialized searching techniques. Here we propose a search technique using blooms filter with a series of n-filters which can significantly increase the search accuracy, with a similar or better storage and communication cost. In this search technique the tradeoff is between storage and false positive rate. This technique is also good against inclusion-relation attacks.

Base Paper:

1. "Fast Phrase Search for Encrypted Cloud Storage" Hoi Ting Poon, Member, IEEE, and Ali Miri, Member, IEEE

2.Introduction:

In the current world everything runs on cloud computing, from saving your personal data to commercial data. Say if you want to save some data onto your personal device, to access the data from somewhere else you need to have direct access to that data, that is not the scenario today. We are not facing the issue due to use of cloud in our everyday life. It even provides platforms to run different architectures depending on the requirement such as infrastructure as a service(IaaS), Platform as a Service(PaaS), Software as a Service(SaaS). Most of us have made cloud a part out everyday life. But when it comes to security, it plays a vital role. As most of us save important documents on email and share sensitive information through it, any breach in security will compromise the data on cloud. Cloud vendors offer encryption at different levels, but the private keys to that encryption are stored at vendors side they are not given to customers. Which means they will be able to access your data at their own will. This is an important concern which has made researchers to work extensively on the topic.

Key functions of cloud include saving and retrieval of data. In current time we use the same set of standard techniques to find data saved on cloud as we used in the earlier days of cloud. There has been less research in the field of searching encrypted data on cloud server. There is some proposed technique to find data over cloud but they have their own drawbacks when we will see in later sections.

3.Existing System:

The former enables the verification of existence of keywords in individual documents, by simply adding the keywords as members, and the latter allow the identification of keyword locations, by concatenating keywords to their locations prior to adding them as members.

The conceptually simple scheme achieves the lowest storage cost among existing solutions. However, its space-efficiency comes at the cost of requiring brute force location verification during phrase search. Since all potential locations of the keywords must be verified, the amount of computation required grows proportionally to the file size.

4.Proposed System:

The proposed a phrase search scheme that achieved further reduction in storage cost. The technique exploits the space-efficiency of Bloom filters to perform conjunctive keyword search and phrase search.

Similar to other techniques a set of keywords to document Bloom filters and a set of keyword location filters are used.

The former enables the verification of existence of keywords in individual documents by simply adding the keywords as members and the latter allow the identification of keyword locations by concatenating keywords to their locations prior to adding them as members.

The conceptually simple scheme achieves the lowest storage cost among existing solutions.

5.Literature review:

1. “Public key encryption with keyword search,” This paper mostly concentrates in search technique using public key schema. Here the schema was implemented in an email scenario where the user can search for a particular keyword from emails and view it without seeing the content of the email. As a result, user can flag certain encrypted emails from the regular emails and take necessary action and the rest of the file will be sorted into their predefined folders.

2. “A secure conjunctive keywords search over encrypted cloud data against inclusion-relation attack,” They investigated the problem for searching over encrypted audit logs. Many of the early works focused on single keyword searches. Recently, researchers have proposed solutions on conjunctive keyword search, which involves multiple keywords.

3. “A low storage phrase search scheme based on bloom filters for encrypted cloud services”, Bloom filters are space-efficient probabilistic data structure used to test whether an element is a member of a set. A Bloom filter contains m bits, where k hash functions, H i (x), are used to map elements to the m -bits in the filter. The Bloom filter is initially set to all zeros. To add an

element, a, to the filter, we compute $H_i(a)$ for $i = 1$ to k and set the corresponding positions in the filter to 1. F While Bloom filters have no false-negatives, it can falsely identify an element as member of a set. Given k hash functions, n items inserted and m bits used in the filter, the probability of false positives is approximately $p = (1 - e^{-kn/m})^k$ and minimum false positive rate is achieved when $k = m/n \ln 2$.

6.Technologies Used:

Java Development Kit 8

Apache Tomcat 9

Mysql Server 5.1

Eclipse Oxygen IDE

HeidiSQL 3.2

Java Development Kit 8:

Java Development Kit is one of the most extensively used programming platforms due to its vast features which support large programming libraries which can be implemented in java. It is owned by Oracle Corporation. JDK provides a Java Runtime Environment and Java Virtual Machine which help in setting up and running java applications. One of the main advantages of using java as a development platform is the cross-platform computing environment that java provides.

Apache Tomcat 9:

Apache Tomcat is a Java Servlet Container when refers to Tomcat server. It is an opensource software designed to run the front-end implementation of Java program on a web browser. It is developed by Apache Software Foundation(ASF) which is non-profit organization to help and support Apache projects. Tomcat has three main components Catalina which a servlet container and implements JavaServer Pages(JSP), Coyote which acts as the connector component to support HTTP protocol and Jasper which is the Tomcats JSP engine that is it compiles java code into servlets.

There has ben many releases of Tomcat of which Apache Tomcat 9 is the latest release. It supports Servlet 4.0, JSP 2.4, EL 3.1 specifications. Tomcat is released under Apache License Version 2.

Mysql 5.1:

Mysql is an open source relational database management system written in C and C++. It is a cross-platform database management system. It is licensed under GPL version 2. Its compatibility with PHP and other scripts which are based on web browsers makes it user friendly to use even on data bases of large size. By default it can accommodate large amounts of data, it has a default size limit of 4GB for tables. We can increase it to our requirements depending on the available memory of the system. Because of the license it is released under, user can make changes to the default code and run it depending on his requirements.

Eclipse Oxygen IDE:

Eclipse is and Integrated Development Environment which is extensively used for java programming. Its basic 2 components are a plug-in and a workspace. Plug-in helps with the adding new features to eclipse environment and Work space it the text editor where the base program is written and compiled. Although most users use eclipse to for developing java applications we can run C, C++, C#, COBAL, JavaScript, Perl, Scala, many more including LaTeX to develop text documents. It is releases as Eclipse Software Development Kit(SDK). It is Released under Eclipse public license which is incompatible with General Public License(GPL). Many Versions of Eclipse has been released, each and every version has its unique name. Current version 4.7 is called Oxygen and the upcoming version which is planned to release in June 2018 is called Photon. The main working of Eclipse is provided by the plug-in on top of the runtime environment. By default, Eclipse can compile and run full features of a java application. Eclipse can be used while installing Apache Tomcat to help in easy deployment of the applications on web.

HeidiSql:

Heidisql is an interface to work with Mysql and its forks. It is an open source software which makes it easy to work from the frontend of the database. Some of its main features include server and client protocols for compatible servers, interface with server, manage users on the server, view or filter variables, kill processes on server, connect to a database and do basic table operations.

Blooms Filter:

Blooms filter is a probabilistic data structure designed to address space problem in search algorithms. It uses false positives to to check if the element is present or not in the given space. It checks if the element is present in the set by checking the probability with either “possibly in set” or “definitely not in set”.

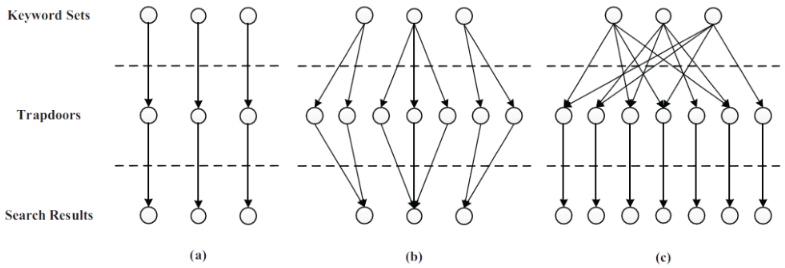

Figure 1. Blooms Filter

Larger the size of the set more the false positives. Blooms filter was initially found to deal with the complex problems of the word search, an example where it addresses search problem is finding a word in dictionary, when finding 90% of the words is easy to find with the conventional methods which depend on memory. The other 10% require a lot memory to find the words. By using blooms filter we can access the data using error-hash could be used to eliminate the unnecessary disk accesses. In more general terms 1 bit of data requires 0.1% false positive rate.

Bloom filter is used here along with a trap door function to access data more securely. The main functionality of trap door is one-way access to data that is, sending data is encrypted and from the receiver's side there will be no means to find who sent the data.

7.Process Flow:

The basic flow of process is while uploading the file is shown in Fig 2, First the user must register as a respective user. Depending on the purpose of the usage he will either be a data owner or a user. A data owner will upload the file which will be encrypted in the process of upload. A data user will search the data uploaded by the data owner in the cloud by using the key provided by the data owner.

To delete data from the cloud the data owner can push a request to server to remover the file which causes all the values related to that particular file to NULL. If the data user wishes to delete a file it is not made available as it might cause a major data breach which causes the data to be used in wrong way.

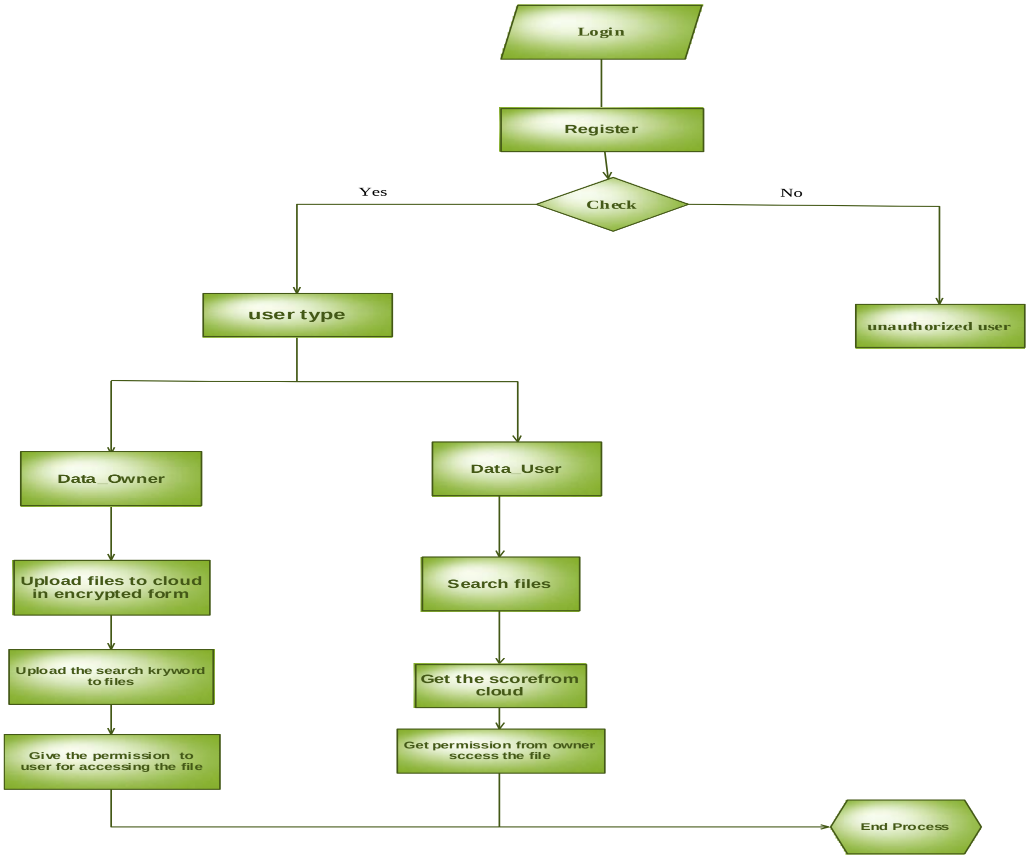
Login
Register
Yes
Check
No
user type
unauthorized user
Data_Owner
Data_User
Upload files to cloud in encrypted form
Search files
Upload the search kryword to files
Get the scorefrom cloud
Give the permission to user for accessing the file
Get permission from owner sccess the file
End Process

Figure 2, Data upload process

8.Future Work:

By using fast phrase search in text documents or on data available on web to find a find keywords and implement natural language techniques to find meaning to sentences which a user gives as input in human language like English.

9.Conclusion:

One of the main advantages of this approach when compared to conventional key word search is the random distribution of the computational load on the server which makes it faster to access data and removes additional unnecessary computational load on the server.

One main drawback of this technique is, if the size of the document is larger with a large number of distinctive words then the size of the filter will be too large and their will be large false positive rate which makes a storage overhead taking large amounts of memory.

10.References:

1. “Fast Phrase Search for Encrypted Cloud Storage” Hoi Ting Poon, Member, IEEE, and Ali Miri, Member, IEEE.

2. D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in In proceedings of Eurocrypt, 2004, pp. 506–522.

3.K. Cai, C. Hong, M. Zhang, D. Feng, and Z. Lv, “A secure conjunctive keywords search over encrypted cloud data against inclusion-relation attack,” in IEEE International Conference on Cloud Computing Technology and Science, 2013, pp. 339–346.

4.” A low storage phrase search scheme based on bloom filters for encrypted cloud services” to appear in IEEE International Conference on Cyber Security and Cloud Computing, 2015.