# Notes

Tradeoff => Efficiency vs Security
Balance can be found in SSE & Order-Revealing Encryption.

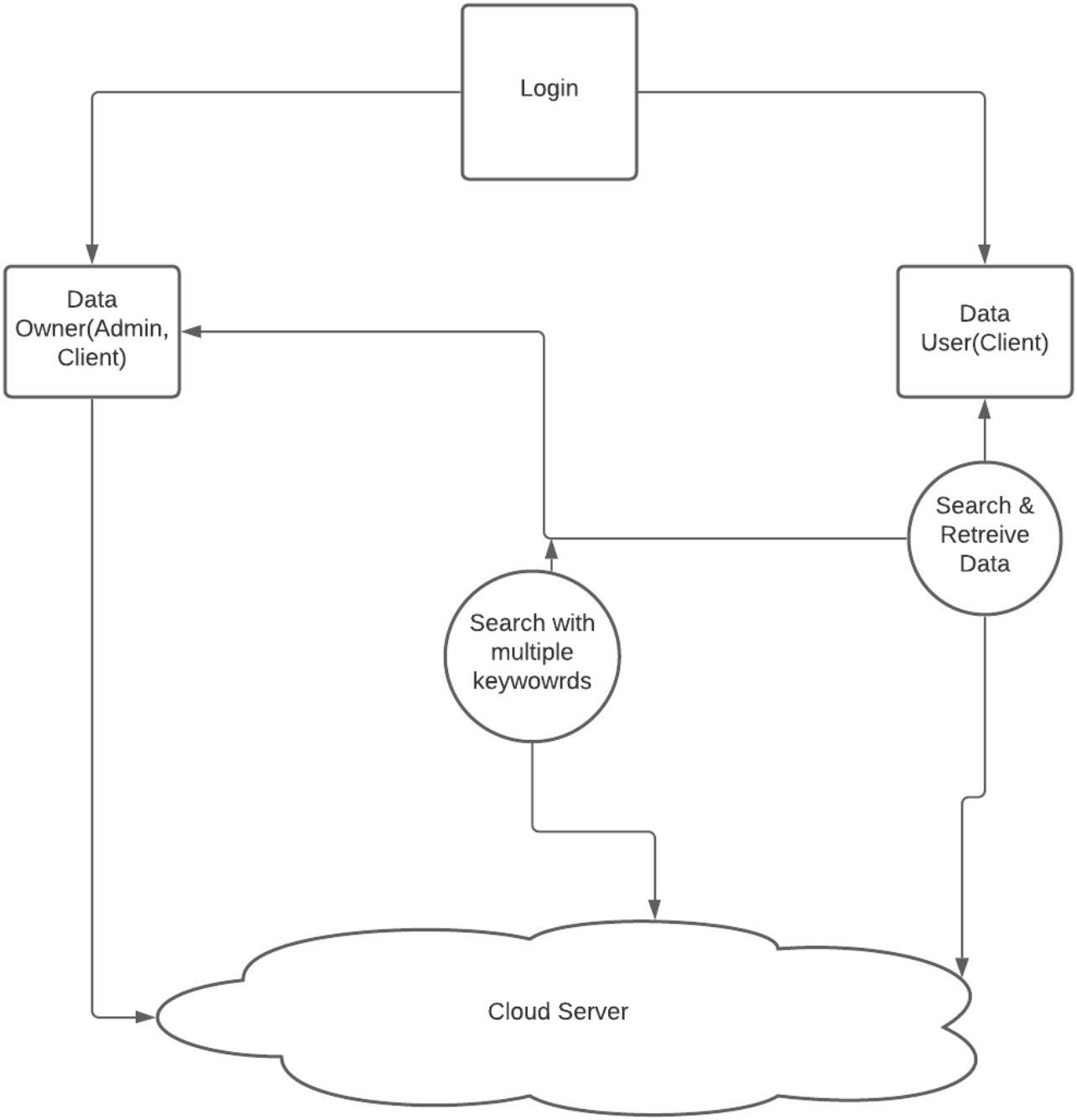
## SSE

Hiding Acess Patterns :

1. Generate false positives and negatives to mask the data.
2. Ensure query tokens are different.
3. We need fresh randomness in query pattern.

**Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing**

FLOW DIAGRAM:




1. Different data owners use different keys to encrypt their files and keywords
2. Authenticated data users can issue a query without knowing secret keys of these different data owners.
3. all of these multi-keyword search schemes can’t support ranked search, which means the cloud server should send back the top-k most relevant files rather than all relevant files. Ranked search is highly desirable in the ‘‘pay-as-you use’’ cloud paradigm, because it can improve the accuracy and efficiency of query.
4. First multi-keyword ranked search scheme (MRSE). Their scheme utilized ‘‘inner product similarity’’ to compute the relevant scores. However, the search time of MRSE is almost **linear** to the number of files in the data set, because we need to compute a relevant score on each file, even though these files don’t contain any searched keyword.

Present Works:

| Name | Techniques Used | Advantages | Disadvantages |
|---|---|---|---|
| multi-keyword ranked search scheme (MRSE) | inner product similarity | | MRSE is almost linear to the number of files in the data set. |
| Sun et al 2013 [Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking] | tree-based ds which exploits vector space model with cosine measure to evaluate the similarity. | Search Efficiency improved | search precision is reduced to some extent. Since the data is encrypted updating the data is difficult.Updating data dynamically is a challenge |
| Fu et al [Achieving effective cloud search services: Multi-keyword ranked search over encrypted cloud data supporting synonym query,2014] | synonym-based multi-keyword ranked search scheme. | more accurate search result and synonym query | Updating cost is excessive. |

## Bloom filter to the rescue!

Advantages:
1.Low storage cost comparing with the existing phrase search schemes.

Disadvantages:
1.Previous schemes cannot achieve multi-keyword ranked search.

## Solve by new approach!

## Paper: Dynamic Multi-Keyword Ranked Search Basedon Bloom Filter Over Encrypted Cloud Data

1.The index tree based on Bloom filter will be designed to improve
the search efficiency.
2. Utilizes vector space model to build an index vector for every file in the outsourcing data set.
3. The cosine similarity measure is used to compute the similarity score of one file to the search query and TF × IDF weight will be used to improve the search accuracy.
4. Our scheme can achieve update operation explicitly and the updating cost of our scheme is low because of the characteristics of the Bloom filter.

**AIM:**

The main aim of this project to retrieve the data owner files based on multi keyword search and calculate term frequency using (TF) × inverse document frequency (IDF)”

**EXISTING SYSTEM:**

· A general approach to save from harm the data confidentiality is to encrypt the data before outsourcing. Searchable encryption schemes allow the client to store the encrypted data to the cloud and execute keyword search over cipher text domain.

· So far, abundant works have been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi-keyword ranked search, etc.

· Show file’s rank based on the keyword search and also ‘n’ number of user’s download the data owner files.

**DISADVANTAGES:**

· Huge cost in terms of data usability.

· For example, the existing techniques on keyword-based information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data.

· Downloading all the data from the cloud and decrypt locally is obviously impractical. Existing System methods not practical due to their high computational overhead for both the cloud sever and user.

**PROPOSED SYSTEM:**

· In our proposed, a secure tree-based search scheme over the encrypted cloud data, which supports multi-keyword ranked search and dynamic operation on the document collection.

· Specifically, in the index building and query generation, the vector space model and the widely used "word frequency (TF)" model are combined to provide multi-keyword ranked search.

· Our system introduces Bloom Filter algorithm based on the index search tree to obtain the sublinear search time.

· And, our systems search effectiveness as well as index tree building effectiveness are better than other associated schemes.

**ALGORITHM USED:**

· Bloom Filter technique

· AES(Attribute Based Encryption)algorithm

. TF/IDF(Term Frequency)

**MODULES:**

In this project, we have three modules

1.Data owner

2.Data user

3..Cloud server

1.Data owner

· Register

· Login

· Upload files

· View files

· Logout

3.Data user

· Register

· Login

· Search files

· Download Files

· Logout

3.Cloud server

· Login

· View data owner

· View data user

· Send key

· Result

· Logout

**MODULES DESCRIPTION:**

1.Data owner

Data owner register their own details. login their account. Upload the files to cloud in encrypted format by using AES(Attribute Encryption Standard) algorithm. At the time of file uploading we calculate(TF/IDF) term frequency and pre processing(stemming,stopwords) for every files. For key generation(Private and Trapdoor key)we are using random key generator and bloom filter technics. Owner can view all files uploaded by own and download the file. Data owner logout their account.

2.Data user

Data user register their own details. user login their account. User can search the files by using keyword and k-value. User can send the request for secret key to cloud for download the files in decrypt format. Using private and trapdoor key users can download the files. Data user logout their account.

3.Cloud server

Cloud server login their account. View data owner and data user details. accept the user key request and send the key user’s registered mail. Finally graph will show based on between number of files and rank of those files, between number of files and request count of those files. Cloud server logout their account.