**IMPLEMENTATION**

### 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.

**MODULE DESCRIPTION:**

# Number of Modules

After careful analysis the system has been identified to have the following modules:

1. **Data Owner**
2. **Encrypted Document Collection**
3. **Cloud Server**
4. **Attribute-based Retrieval Feature and ARF Tree**

# MODULES DESCRIPTION:

**Data Owner**

The data owner is responsible for collecting and preprocessing the documents, and then obtains a set of high quality files F. He sets the attributes for each document and then hierarchically encrypts the document collection based on attributes. In addition, an index vector is extracted from each document based on the document’s content and attributes. An index structure I is constructed based on the index vectors of the documents. At last, both the encrypted documents C and encrypted index structure are sent to the cloud server. The cloud server is responsible for storing the encrypted documents and executing document search based on the index structure.

**Encrypted Document Collection**

The efficient retrieval scheme over encrypted document collection is designed and we first describe the process of constructing the ARF tree. Then a depth-first searching algorithm of the ARF tree is designed and in addition, it can be operated in a parallel manner flexibly. Given a collection of documents F = {F1; F2; · · · ; FN}, each document needs to be scanned for one time and the number of each keyword is recorded. Then a normalized vector for the document is generated based on the keyword dictionary. The attribute vector of a document can be built based on attribute dictionary A and the associated attributes assigned by the data owner. Organizing the document vectors properly can significantly improve the search efficiency. In some encrypted document retrieval schemes, the document content vectors are organized randomly and the search complexity is O(N), where N is the number of documents. To improve search efficiency, in some other schemes, the vectors are organized based on their relative similarities and they can obtain sublinear search efficiency.

**Cloud Server**

The Cloud Server and CA center are assumed to be trustable. we focus on the security of the proposed hierarchical document encryption scheme and its security mainly involves two aspects including document confidentiality and content keys confidentiality. The documents are encrypted based on symmetric encryption schemes (e.g., AES) with content keys and their security is out of the scope in this paper. In this section, we analyze the security of the content keys which are encrypted by the proposed hierarchical encryption scheme. We provide theDecisional Bilinear Diffie-Hellman assumption.

**Attribute-based Retrieval Feature and ARF Tree**

To improve the search efficiency of multi-keywords search process, a height- balanced index tree named ARF tree is built based on the document vectors. Specifically, the document vectors are organized as clusters according to their similarities. Each node in the tree represents a cluster composed of a set of document vectors or sub-clusters. An ARF vector is a quintuple summarization about a cluster. Given P documents {Fj} where j = 1; 2; · · · ; P, we assume that a cluster C comprises the document vectors of {Fj}, i.e., {Vj ; V ′ j } where j = 1; 2; · · · ; P. Then, the ARF vector of the cluster is defined as follows: ARF = (P; LS; SS; Vmax;Amin), where P is the number of document content vectors in the cluster, LS is the linear sum of the P content vectors, i.e., LS = ΣP j=1 Vj ,

SΣS is the square sum of the P content vectors, i.e., SS = P j=1 V 2

j , Vmax denotes a vector consisting of m values which are calculated