**Fast Phrase Search for Encrypted Cloud Storage**

**ABSTRACT:**

Cloud computing has generated much interest in the research community in recent years for its many advantages, but has also raise security and privacy concerns. The storage and access of confidential documents have been identified as one of the central problems in the area. In particular, many researchers investigated solutions to search over encrypted documents stored on remote cloud servers. While many schemes have been proposed to perform conjunctive keyword search, less attention has been noted on more specialized searching techniques. In this paper, we present a phrase search technique based on Bloom filters that is significantly faster than existing solutions, with similar or better storage and communication cost. Our technique uses a series of n-gram filters to support the functionality. The scheme exhibits a trade-off between storage and false positive rate, and is adaptable to defend against inclusion-relation attacks. A design approach based on an application’s target false positive rate is also described.

**EXISTING SYSTEM:**

* Boneh et al. proposed one of the earliest works on keyword searching. Their scheme uses public key encryption to allow keywords to be searchable without revealing data content.
* Waters et al. 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.
* Other interesting problems, such as the ranking of search results and searching with keywords that might contain errors termed fuzzy keyword search, have also been considered. The ability to search for phrases was also recently investigated.
* Some of the existing system has examined the security of the proposed solutions and, where flaws were found, solutions were proposed

**DISADVANTAGES OF EXISTING SYSTEM:**

* The cloud can read any data it desired, providing no privacy to its users. The storage of private keys and encrypted data by the cloud provider is also problematic in case of data breach.
* By recognizing the almost exponential distribution of keywords, the entries in the keyword location tables are split into pairs to achieve normalization without the high cost of storing unused random data. However, the use of encrypted indexes and the need to perform client-side encryption and decryption may still be computationally expensive in certain applications.
* Its space-efficiency comes at the cost of requiring a 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. As a result, the scheme exhibits a high processing time.

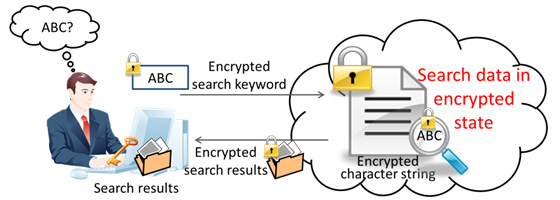
**PROPOSED SYSTEM:**

* In this paper, we present a phrase search scheme which achieves a much faster response time than existing solutions. The scheme is also scalable, where documents can easily be removed and added to the corpus. We also describe modifications to the scheme to lower storage cost at a small cost in response time and to defend against cloud providers with statistical knowledge on stored data.
* Although phrase searches are processed independently using our technique, they are typically a specialized function in a keyword search scheme, where the primary function is to provide conjunctive keyword searches. Therefore, we describe both the basic conjunctive keyword search algorithm and the basic phrase search algorithm along with design techniques.

**ADVANTAGES OF PROPOSED SYSTEM:**

* Our framework differs from some of the earlier works, where keywords generally consist of meta-data rather than content of the files and where a trusted key escrow authority is used due to the use of Identity based encryption.
* When compared to recent works, where an organization wishes to outsource computing resources to a cloud storage provider and enable search for its employees, where the aim is to return properly ranked files. Most other recent works related to search over encrypted data have considered similar models such as, where the client acts as both data owner and user.

**SYSTEM ARCHITECTURE:**



**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium Dual Core.
* Hard Disk : 120 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 1 GB

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 7.
* Coding Language : JAVA/J2EE
* Tool : Netbeans 7.2.1
* Database : MYSQL