**Optimizing Information Leakage in Multicloud Storage Services**

**ABSTRACT:**

Many schemes have been recently advanced for storing data on multiple clouds. Distributing data over different cloud storage providers (CSPs) automatically provides users with a certain degree of information leakage control, for no single point of attack can leak all the information. However, unplanned distribution of data chunks can lead to high information disclosure even while using multiple clouds. In this paper, we study an important information leakage problem caused by unplanned data distribution in multicloud storage services. Then, we present StoreSim, an information leakage aware storage system in multicloud. StoreSim aims to store syntactically similar data on the same cloud, thus minimizing the user’s information leakage across multiple clouds. We design an approximate algorithm to efficiently generate similarity-preserving signatures for data chunks based on MinHash and Bloom filter, and also design a function to compute the information leakage based on these signatures. Next, we present an effective storage plan generation algorithm based on clustering for distributing data chunks with minimal information leakage across multiple clouds. Finally, we evaluate our scheme using two real datasets from Wikipedia and GitHub. We show that our scheme can reduce the information leakage by up to 60% compared to unplanned placement. Furthermore, our analysis on system attackability demonstrates that our scheme makes attacks on information more complex.

**EXISTING SYSTEM:**

* Li et al. proposed a privacy loss measure based on the JS-divergence distance which is a method of measuring the similarity between two probability distributions.

**DISADVANTAGES OF EXISTING SYSTEM:**

* Storing the data in a multicloud system without proper optimization on the data distribution can lead to avoidable information leakage.
* Their work failed to reveal optimization aspects of information leakages of the commercial CSPs they studied.

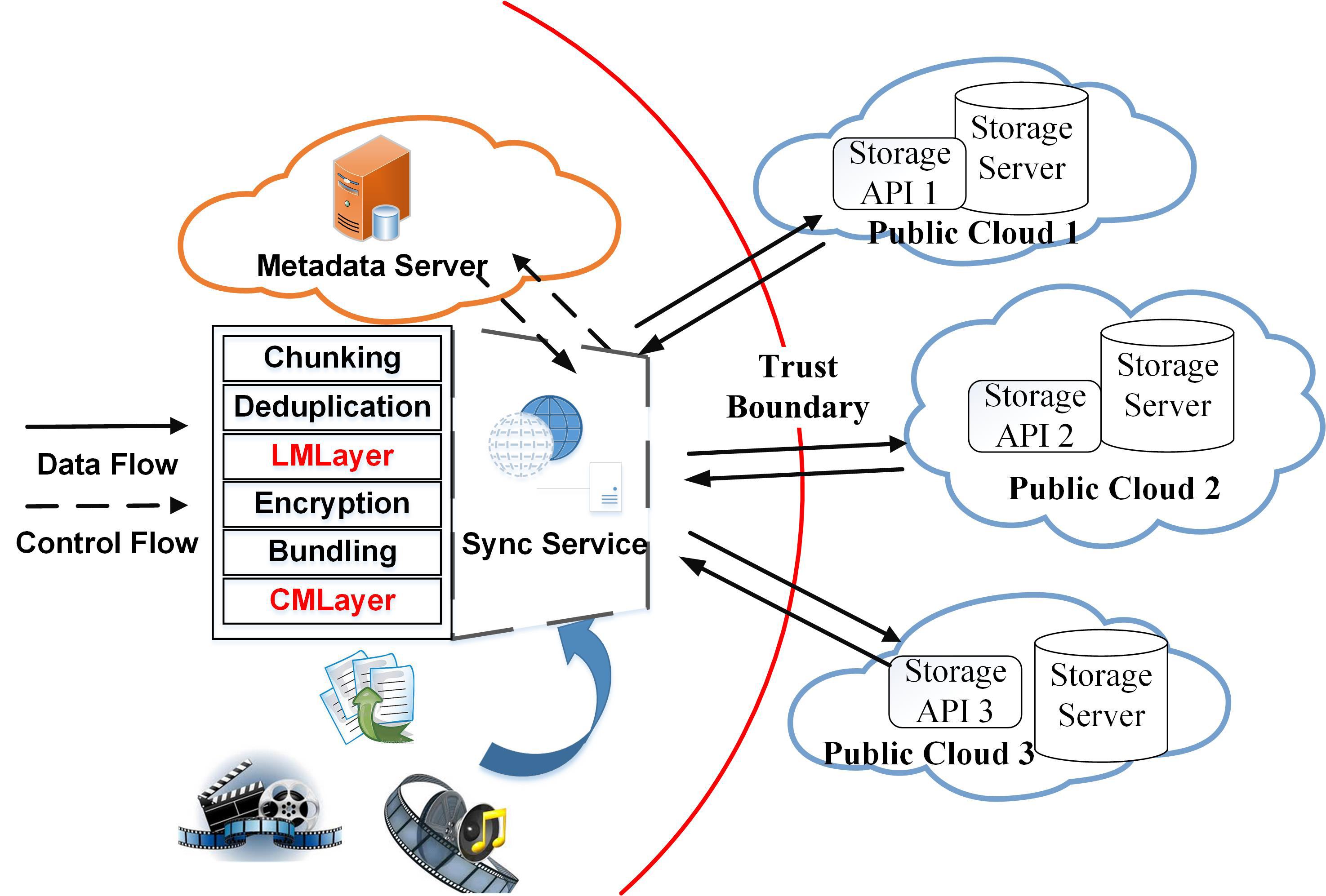
**PROPOSED SYSTEM:**

* In this paper, we focus on reducing information leakage to each individual CSP in a multicloud storage system and provide mechanisms for distributing users data over multiple CSPs in a leakage aware manner. First we provide a novel algorithm for generating similarity preserving signatures for data chunks. Next based on this algorithm, we devise a chunk placement storage plan that efficiently synchronizes similar chunks together in a multicloud environment.
* We present StoreSim, an information leakage aware multicloud storage system which incorporates three important distributed entities and we also formulate information leakage optimization problem in multicloud.
* We propose an approximate algorithm, BFSMinHash, based on Minhash and Bloom filter to generate similarity-preserving signatures for data chunks. We also design a pairwise information leakage function based on Jaccard similarity.

**ADVANTAGES OF PROPOSED SYSTEM:**

* We show the effectiveness and efficiency of our proposed scheme for reducing information leakage across multiple clouds. Furthermore, our analysis on the system attackability demonstrates that StoreSim makes attacks on information much more complex.
* To the best of our knowledge, this is the first work which applies near-duplicate techniques for preventing information leakage in multicloud storage services. Our work focuses on the information leakage optimization for storage service in a multicloud environment by exploiting information similarity caused by the synchronization of modified data.

**SYSTEM ARCHITECTURE:**



**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium i3 Processor
* Hard Disk : 500 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 2 GB

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 10.
* Coding Language : JAVA.
* Tool : Netbeans 8.2
* Database : MYSQL

**REFERENCE:**

Hao Zhuang, Member, IEEE, Rameez Rahman,Pan Hui, Member, IEEE, and Karl Aberer, Member, IEEE, “Optimizing Information Leakage in Multi-cloud Storage Services”, IEEE 2020.