**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.

**SYSTEM ANALYSIS**

**Existing System:**

* In fact, the data deduplication technique, which is widely adopted by current cloud storage services in existing clouds, is one example of exploiting the similarities among different data chunks to save disk space and avoid data retransmission . It identifies the same data chunks by their fingerprints which are generated by fingerprinting algorithms such as SHA-1, MD5. Any change to the data will produce a very different fingerprint with high probability . However, these fingerprints can only detect whether or not the data nodes are duplicate, which is only good for exact equality testing. Determining identical chunks is relatively straightforward but efficiently determining similarity between chunks is an intricate task due to the lack of similarity preserving fingerprints (or signatures).

**Disadvantages:**

* Unplanned distribution of data chunks can lead to high information disclosure even while using multiple clouds.
* Frequent modifications of files by users result in large amount of similar chunks1;
* Similar chunks across files, due to which existing CSPs use the data de duplication technique.

**Proposed System:**

* 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 to generate similarity-preserving signatures for data chunks.
* Based on the information match measured by BFSMinHash, we develop an efficient storage plan generation algorithm, Clustering, for distributing users data to different clouds.

**Advantages:**

* However, previous works employed only a single cloud which has both compute and storage capacity. Our work is different since we consider a mutli cloud in which each storage cloud is only served as storage without the ability to compute.
* Our work is not alone in storing data with the adoption of multiple CSPs these work focused on different issues such as cost optimization , data consistency and availability.

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

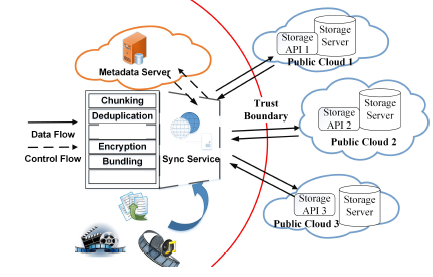
* System : Pentium Dual Core.
* Hard Disk : 120 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 1GB.

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 7.
* Coding Language : JAVA/J2EE
* Tool : Netbeans 7.4
* Database : MYSQL

**SYSTEM DESIGN**

**ARCHITECTURE**

****

**REFENRENCES**

* J. Crowcroft, “On the duality of resilience and privacy,” in Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 471, no. 2175. The Royal Society, 2015, p. 20140862.
* Bessani, M. Correia, B. Quaresma, F. Andr´e, and P. Sousa, “Depsky: dependable and secure storage in a cloud-of-clouds,” ACM Transactions on Storage (TOS), vol. 9, no. 4, p. 12, 2013.
* H. Chen, Y. Hu, P. Lee, and Y. Tang, “Nccloud: A network-coding-based storage system in a cloud-of-clouds,” 2013.
* T. G. Papaioannou, N. Bonvin, and K. Aberer, “Scalia: an adaptive scheme for efficient multi-cloud storage,” in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis. IEEE Computer Society Press, 2012, p. 20.
* Z. Wu, M. Butkiewicz, D. Perkins, E. Katz-Bassett, and H. V. Madhyastha, “Spanstore: Cost-effective geo-replicated storage spanning multiple cloud services,” in Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles. ACM, 2013, pp. 292–308.
* G. Greenwald and E. MacAskill, “Nsa prism program taps in to user data of apple, google and others,” The Guardian, vol. 7, no. 6, pp. 1–43, 2013.
* T. Suel and N. Memon, “Algorithms for delta compression and remote file synchronization,” 2002.
* [8] I. Drago, E. Bocchi, M. Mellia, H. Slatman, and A. Pras, “Benchmarking personal cloud storage,” in Proceedings of the 2013 conference on Internet
* measurement conference. ACM, 2013, pp. 205–212.
* [9] I. Drago, M. Mellia, M.MMunafo, A. Sperotto, R. Sadre, and A. Pras, “Inside dropbox: understanding personal cloud storage services,” in Proceedings of the 2012 ACM conference on Internet measurement conference. ACM, 2012, pp. 481–494.
* [10] U. Manber et al., “Finding similar files in a large file system.” in Usenix Winter, vol. 94, 1994, pp. 1–10.
* [11] P. Mahajan, S. Setty, S. Lee, A. Clement, L. Alvisi, M. Dahlin, and M.Walfish, “Depot: Cloud storage with minimal trust,” ACM Transactions on Computer Systems (TOCS), vol. 29, no. 4, p. 12, 2011.
* [12] A. J. Feldman, W. P. Zeller, M. J. Freedman, and E. W. Felten, “Sporc: Group collaboration using untrusted cloud resources.” in OSDI, vol. 10, 2010, pp. 337–350.
* [13] F. Dabek, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica, “Wide-area cooperative storage with cfs,” in ACM SIGOPS Operating Systems Review, vol. 35, no. 5. ACM, 2001, pp. 202–215.
* [14] L. P. Cox and B. D. Noble, “Samsara: Honor among thieves in peer-to-peer storage,” ACM SIGOPS Operating Systems Review, vol. 37, no. 5, pp. 120–132, 2003.
* [15] H. Zhuang, R. Rahman, and K. Aberer, “Decentralizing the cloud: How can small data centers cooperate?” in Peer-to-Peer Computing (P2P), 14-th IEEE International Conference on. Ieee, 2014, pp. 1–10.
* [16] H. Zhuang, R. Rahman, P. Hui, and K. Aberer, “Storesim: Optimizing information leakage in multicloud storage services,” in Cloud Computing
* Technology and Science (CloudCom), 2015 IEEE 7th International Conference on. IEEE, 2015, pp. 379–386.
* [17] S. Choy, B. Wong, G. Simon, and C. Rosenberg, “A hybrid edge-cloud architecture for reducing on-demand gaming latency,” Multimedia Systems,
* pp. 1–17, 2014.
* [18] T. Zou, R. Le Bras, M. V. Salles, A. Demers, and J. Gehrke, “Cloudia: a deployment advisor for public clouds,” in Proceedings of the VLDB Endowment, vol. 6, no. 2. VLDB Endowment, 2012, pp. 121–132.
* [19] J.-M. Bohli, N. Gruschka, M. Jensen, L. L. Iacono, and N. Marnau, “Security and privacy-enhancing multicloud architectures,” Dependable and Secure Computing, IEEE Transactions on, vol. 10, no. 4, pp. 212–224, 2013.
* [20] H. Harkous, R. Rahman, and K. Aberer, “C3p: Context-aware crowdsourced cloud privacy,” in 14th Privacy Enhancing Technologies Symposium (PETS 2014), 2014.